

**PROYECTO DE
ASISTENCIA
TÉCNICA PARA
ALIMENTACIÓN
Y NUTRICIÓN**

Guía de Muestreo

Robert Magnani

*Apoyo a programas integrados de seguridad alimentaria y
nutrición para mejorar la salud y bienestar
de las mujeres y los niños.*



traducción al español de esta publicación fue realizada gracias al apoyo brindado al proyecto de Asistencia Técnica en Alimentación y Nutrición (FANTA) de la Oficina de Salud y Nutrición, Buró para Programas Globales, Investigación y Apoyo Exterior, de la Agencia para el Desarrollo Internacional de Estados Unidos de América (USAID) bajo los términos del Acuerdo Cooperativo No. HRN-A-00-98-00046-00 adjudicado a la Academia para el Desarrollo Educativo (AED). Las opiniones aquí expresadas pertenecen a los autores y no necesariamente reflejan los puntos de vista de USAID ni de AED. Las versiones anteriores en inglés de la presente guía se prepararon con fondos del Proyecto de Seguimiento de la Alimentación y Nutrición (IMPACT) (Contrato No. DAN-5110-Q-00-0014-00, Orden de entrega 16), administrado por el International Science and Technology Institute, Inc. (ISTI). Las opiniones aquí expresadas son de los autores y no representan necesariamente la política de la Agencia de los Estados Unidos para el Desarrollo Internacional.

Publicado en junio de 2001

Dirección para pedidos de ejemplares de la presente Guía:

Food & Nutrition Technical Assistance Project
(FANTA)
Academy for Educational Development
1825 Connecticut Avenue, NW
Washington, D.C. 20009-5721
Tel.: 202-884 8000. Fax: 202-884 8432
Correo electrónico: fanta@aed.org
Internet: www.fantaproject.org

La

Índice

1. Finalidad de la guía
2. Definición de los objetivos de la medición en el diseño de la muestra
3. Determinación de los requisitos del tamaño de la muestra.....
4. Selección de la muestra
5. Análisis de los datos.....

Recuadro

1. Información sobre esta serie

Figuras

- 3-1. Ejemplo ilustrativo de las necesidades de información para determinar el tamaño de la muestra, con indicadores genéricos de alimentación del lactante y del niño pequeño en programas del Título II
- 3-2. Valores de Z_{α} y Z_{β}
- 3-3. Ejemplo ilustrativo de los cálculos del tamaño de la muestra para indicadores expresados como proporción.....
- 3-4. Tamaño de la muestra necesario para combinaciones selectas de P_1 y cambios o diferencias por detectarse entre los grupos de comparación ($\alpha = 0,95$ y $\beta = 0,80$).....
- 3-5. Ejemplo ilustrativo del cálculo del tamaño de la muestra para indicadores expresados como media
- 3-6. Número típico de familias por entrevistar en búsqueda de una persona de referencia para los indicadores genéricos de salud en programas del Título II (con la suposición de que hay seis personas por familia)
- 3-7. Ejemplo ilustrativo del cálculo del tamaño de la muestra de la encuesta de seguimiento para indicadores expresados como proporción
- 3-8. Ejemplo ilustrativo del cálculo del tamaño de la muestra de la encuesta de seguimiento para indicadores expresados como media
- 4-1. Pasos para seleccionar una muestra aleatoria sistemática de conglomerados con PPS
- 4-2. Ejemplo ilustrativo: selección de una muestra aleatoria sistemática de conglomerados con PPS...
- 4-3. Pasos para seleccionar una muestra aleatoria sistemática de conglomerados con igualdad de probabilidades
- 4-4. Ejemplo ilustrativo: selección de una muestra aleatoria sistemática de conglomerados con igualdad de probabilidades
- 4-5. Pasos para usar el método de segmentación para escoger las familias de muestra.....

- 4-6. Ejemplo de un conglomerado hipotético dividido en seis segmentos.....
- 4-7. Mapa de un conglomerado hipotético de muestra con los posibles puntos de partida
- 4-8. Ejemplo ilustrativo del diseño y la selección de muestras en una zona de comparación.....
- 5-1. Procedimientos para calcular las probabilidades de muestreo de los elementos de la muestra (Pi) con diseños selectos de muestreo por conglomerados en dos etapas
- 5-2. Ejemplo ilustrativo del cómputo de las probabilidades de selección, los valores de ponderación del muestreo y los valores de ponderación normalizados del muestreo: datos hipotéticos.

Apéndices

- Apéndice 1.** Lista de indicadores genéricos de programas del Título II
- Apéndice 2.** Tamaño de la muestra necesario para determinadas combinaciones de P_1 y cambios o diferencias por detectarse entre los grupos de comparación ($\alpha = 0,95$ y $\beta = 0,90$).....
- Apéndice 3.** ¿Cuáles son la confianza y el poder estadísticos de los subgrupos?.....

Nota de agradecimiento

El original en inglés de la presente guía fue redactado por Robert Magnani de la Facultad de Salud Pública y Medicina Tropical de la Universidad de Tulane para el proyecto IMPACT. El autor desea expresar sus agradecimientos a Janet Rice de la Universidad de Tulane por sus valiosos comentarios sobre las versiones preliminares. Eunyong Chung de la Oficina de Salud y Nutrición del Departamento de Programas Mundiales de la USAID proporcionó valiosa información y apoyo para la preparación de la guía. La Oficina de Alimentos para la Paz tuvo una intervención decisiva en el fomento y apoyo de este trabajo. Bruce Cogill, Anne Swindale y Patrick Diskin del proyecto IMPACT ofrecieron extensas observaciones y asistencia. El autor hace llegar un reconocimiento especial a Dorothy B. Wexler, revisora del original en inglés, y a Stacy Swartwood, asesora en diseño. Las instituciones patrocinadoras de actividades realizadas en cooperación desempeñaron una función esencial en la preparación de esta obra. El autor se la dedica a ellas.

1

Finalidad de la guía

Recuadro 1: Información sobre esta serie...

Esta serie de guías de medición de indicadores genéricos para programas del Título II ha sido preparada por el Proyecto de Asistencia Técnica para Alimentación y Nutrición (FANTA) y sus predecesores (IMPACT, LINKAGES), como parte del apoyo prestado por la USAID a las instituciones patrocinadoras cooperadoras en la creación de sistemas de seguimiento y evaluación para empleo en programas del Título II. Estas guías tienen por fin ofrecer la base técnica para los indicadores y el método recomendado para acopiar, analizar y notificar información sobre los indicadores genéricos establecidos en consulta con organizaciones voluntarias privadas en 1995/1996.

La siguiente es la lista de guías disponibles:

Agricultural Productivity Indicators Measurement Guide

Anthropometric Indicators Measurement Guide

Credit With Education: A Promising Title II Microfinance Strategy

Food for Education Indicator Guide

Food Security Indicators and Framework for Use in the Monitoring and Evaluation of Food Aid Programs

HIV/AIDS: A Guide for Nutrition, Care and Support

Infant and Child Feeding Indicators Measurement Guide

Measuring Household Food Consumption: A Technical Guide

Sampling Guide (Guía de Muestreo)

Water and Sanitation Indicators Measurement Guide

La finalidad del muestreo es reducir el costo de la recolección de datos de una población mediante el acopio de información de un subgrupo en lugar de todo el conjunto. A menudo, las encuestas de muestras son el medio más factible de recopilar los datos necesarios para la evaluación de programas del Título II. Esta guía muestra cómo escoger muestras de comunidades, familias o personas para esas encuestas de manera que, cuando se combinen con indicadores y diseños apropiados para estudios de

evaluación, permitan derivar conclusiones válidas sobre la eficacia de los programas del Título II. En la guía se recalca el uso de los métodos de muestreo probabilístico, considerados esenciales para asegurar la objetividad en la evaluación de programas. Puede esperarse que las estimaciones de las características de la población derivadas de encuestas de muestras realizadas siguiendo las directrices recomendadas se aproximen al valor “verdadero” de la población dentro de un margen de error especificado con una probabilidad conocida.

La guía se ha redactado para lectores con limitados conocimientos de muestreo. Sin embargo, el conocimiento de estadísticas básicas será útil para usarla. Los materiales se presentan paso a paso en el orden que podría seguirse al realizar una evaluación de un programa del Título II. Se describen cuatro fases principales:

1. Definición de los objetivos de la medición en el diseño de la encuesta. En esta fase se aborda lo que se espera lograr con la encuesta propiamente dicha. Entraña tanto la sustancia del estudio, es decir, el progreso alcanzado por una población destinataria en el logro de los objetivos del proyecto, como la cuestión estadística del grado de precisión necesario de los datos.

2. Determinación de los requisitos del tamaño de la muestra. En esta fase se explica cómo calcular el tamaño de la muestra después de decidir *qué* debe medirse y *con qué precisión*. El procedimiento se divide en tres pasos principales. Primero hay que definir el número total de elementos de muestra; para eso se proporcionan fórmulas para determinar cuántas personas deben formar parte de la muestra, ya sea que el progreso se mida por los cambios en la *proporción* de la población con una característica determinada o por los cambios en la *media* de un indicador dado (por ejemplo, el consumo total de calorías diarias per cápita). Segundo, el número total de elementos debe convertirse en el número de familias con las que se debe establecer contacto. Tercero, el número total de familias necesita convertirse en unidades de uso práctico (conglomerados y sujetos integrantes de los mismos) que serán visitadas por el equipo de encuesta.

3. Selección de la muestra. En esta fase se define el muestreo probabilístico y se explica por qué se recomienda. Luego se explican, paso a paso, varias formas de selección de los conglomerados y elementos citados, según las circunstancias (particularmente si se sabe o no cuál es el tamaño del conglomerado). También se ofrecen sugerencias sobre la forma de abordar los problemas operativos.

4. Análisis de los datos. En esta fase se abordan las cuestiones estadísticas relacionadas con el cálculo de los valores de ponderación y los errores estándar surgidos como resultado de la combinación de los métodos de selección de los conglomerados y elementos. Se ofrecen fórmulas para calcular los valores de ponderación correspondientes a varias combinaciones típicas.

Esta obra difiere de otras guías de muestreo destinadas al personal asignado al terreno en el sentido de que ofrece breves explicaciones de la razón fundamental de varios procedimientos y prácticas de muestreo. Se cree que el personal asignado al terreno estará mejor preparado para adaptar los

procedimientos a las circunstancias locales si entiende la razón fundamental de un procedimiento dado. En toda la guía se ofrecen ejemplos ilustrativos de los cálculos y procedimientos correspondientes.

Aunque la guía se ha redactado para abordar cuestiones de muestreo que pueden surgir durante las evaluaciones de los programas del Título II, en ningún documento de este tipo se pueden prever a cabalidad todos los posibles matices de la aplicación real. Por eso, los usuarios deben suponer que, de vez en cuando, necesitarán consultar a especialistas en muestreo.

2

Definición de los objetivos de la medición en el diseño de la encuesta

El primer paso para diseñar una encuesta es la definición de los objetivos de la medición. Esto reviste particular importancia cuando la encuesta debe proporcionar los datos primarios para la evaluación de los programas. Cuando se especifican claramente los objetivos, se pueden incluir preguntas apropiadas en los protocolos de la encuesta y diseñar un plan de muestreo apropiado para darles cabida. Cuando la definición es inadecuada, tanto la encuesta como la evaluación del programa pueden desviarse de su curso desde el principio.

La definición de los objetivos de la medición consiste en responder a las tres preguntas siguientes:

- ¿Qué se debe medir?
- ¿En quién?
- ¿Con qué grado de precisión?

¿Qué se debe medir?

La pregunta ¿“qué se debe medir”? suele responderse con *variables* o *indicadores*.¹ Los indicadores recomendados para varios programas del Título II se presentan en otras guías del proyecto IMPACT y, por esa razón, no se estudiarán a fondo aquí. Los recomendados para las evaluaciones se refieren a los tipos de resultados o efectos que se pretende producir en el programa: por ejemplo, reducción del número de niños de 6 a 59 meses con retraso del crecimiento o aumento del número de niños menores de 6 meses alimentados exclusivamente con leche materna.

Otros dos aspectos de la pregunta ¿“qué se debe medir”? merecen atención: (1) si se deben medir los cambios con el tiempo o en el grupo destinatario en comparación con el grupo o zona testigo y (2) qué factores de confusión pueden surgir cuando se emplea un grupo o zona testigo.

1. En una situación ideal, lo que debe medirse se definiría con la preparación de “cuadros ficticios” para el informe final. Esos cuadros indicarían la forma de emplear en el análisis cada dato medido en una encuesta dada.

La importancia del primer asunto (si se deben medir los cambios con el tiempo o las diferencias entre los grupos del proyecto y los testigos) radica en que los requisitos del tamaño de la muestra varían mucho, según lo que se seleccione (véase una explicación más detallada en el capítulo 3). La importancia del segundo radica en la necesidad de identificar desde el principio cualquier factor de confusión que pudiera sesgar los resultados en el análisis de datos de grupos o zonas testigos. Naturalmente, hay que hacer todo lo posible por escoger una zona testigo que sea lo más parecida posible a la zona del programa. Sin embargo, como es casi inevitable que haya algunas diferencias, es preciso especificar cualquier factor o variable que se crea que puede influir en los indicadores de los resultados de la evaluación. Eso permitirá medirlos en los protocolos de la encuesta. Ese asunto se discute con más detalles en la Guía de seguimiento y evaluación del proyecto IMPACT.

¿En quién?

La pregunta *¿en quién?* ofrece la base para definir (1) la población a la cual se pueden extrapolar válidamente los resultados de la encuesta y (2) la amplitud del muestreo y de las operaciones que deben realizarse en el trabajo sobre el terreno.

En este caso hay que abordar varios asuntos.

Dominios

Un dominio es una población o un subgrupo específico en el que se desea hacer estimaciones separadas en la encuesta. Para las evaluaciones de proyectos del Título II, los dominios constarán normalmente de (1) la población en general de la zona destinataria del proyecto o (2) la subpoblación de beneficiarios del proyecto. Cuando se usan grupos o zonas testigos, esa subpoblación constituirá un dominio adicional.

Los dominios necesitan definirse al comienzo puesto que los requisitos del tamaño de la muestra se determinan sobre la base del dominio (véase el capítulo 3). Por ejemplo, si se pretende comprar las nuevas zonas del proyecto con las antiguas, esos dos grupos tendrían que identificarse como dominios separados. Eso asegurará que el tamaño de la muestra en cada grupo sea suficiente para hacer comparaciones significativas. El proceso de dividir a la población objeto de estudio en subgrupos o dominios separados se llama *estratificación*.

La designación de una población o un subgrupo particular de interés como dominio es la única forma de asegurarse de que el tamaño de la muestra sea suficiente para medir con seguridad los cambios con el tiempo en el subgrupo o las diferencias entre los grupos de comparación.

Universo de la encuesta

El universo se refiere a la población o a la región geográfica sobre la que se pueden hacer inferencias a partir de los datos de la encuesta. En esta guía, el universo será normalmente la población de la región geográfica cubierta por el proyecto objeto de evaluación. Si hay un grupo testigo, su universo sería la población de no beneficiarios definida por región geográfica de la cual se ha escogido la muestra testigo. Si faltan recursos para cubrir todo el universo de la encuesta, por ejemplo, se podría emplear un universo parcial de tres de los cinco distritos cubiertos por un proyecto. Sin embargo, esto limitaría la capacidad de generalizar los datos de la encuesta (y los resultados de la evaluación) a un universo más pequeño.²

Unidades de medición y entrevistados

Las unidades de medición son las personas *a las cuales* se refieren los datos de la encuesta y los entrevistados son las personas *de las cuales* se obtiene la información. Por lo general, las unidades de medición se definen en los indicadores. De ordinario, los indicadores para los programas del Título II se refieren a las familias. La excepción está constituida por los indicadores de la alimentación del lactante en que los lactantes y otros niños menores de 24 meses son las unidades. Los entrevistados y las unidades de medición pueden ser iguales o no. Son iguales cuando corresponden a los indicadores para los cuales se pide a los entrevistados que den información sobre sí mismos. Difieren cuando la información se obtiene de lo que se llama entrevistados *sustitutivos*. Por ejemplo, la información para los indicadores de la alimentación del lactante se obtiene típicamente de la madre o del dispensador de cuidado de cada niño. Los indicadores referentes a la familia suelen medirse con la información obtenida al entrevistar a una persona de la familia con buenos conocimientos.

Es preciso identificar las unidades de medición y los entrevistados en relación con cada indicador desde el comienzo del proceso de diseño de la encuesta porque eso influirá en la elaboración del plan de muestreo y en la calidad de los datos acopiados en la encuesta.

¿Con qué grado de precisión?

-
2. Se aplica la regla de la capacidad limitada para generalizar aun si se escogen los distritos al azar. Eso sucede porque la aleatorización exige un número suficientemente grande de unidades de muestreo para producir estimaciones sin ningún sesgo. Se podría defender el uso de una muestra de distritos para “representar” a un conjunto mayor, si se pudiera demostrar empíricamente que los distritos de la muestra son similares en cuanto a las principales características y que la ejecución del programa no favorece a los distritos escogidos para la evaluación.

El nivel o el grado de *precisión* necesario para una encuesta se refiere a la magnitud del error en las estimaciones de la encuesta que se considere tolerable para una tarea en particular. Las encuestas pueden diseñarse para proporcionar estimaciones muy precisas o solamente aproximaciones generales. El grado de precisión o de aproximación varía según los recursos disponibles y el uso que se pretenda dar a los datos de la encuesta.

En las encuestas destinadas a medir el cambio con el tiempo o las diferencias entre los grupos de comparación, la precisión se especifica como el cambio *mínimo* o la diferencia *mínima* entre los grupos de comparación que se puede medir con seguridad. Esa será una decisión del diseñador de la encuesta, según el grado de significación

En las encuestas destinadas a medir el cambio con el tiempo o las diferencias entre los grupos de comparación, la precisión se especifica como el cambio mínimo o la diferencia mínima entre los grupos de comparación que se puede medir con seguridad.

estadística y el poder estadístico deseados (véase el capítulo 3, particularmente las secciones 1 y 3.6). También dependerá de las metas del programa que se hayan especificado para un indicador particular.

La tarea de especificar los requisitos de precisión de la encuesta para las evaluaciones de los programas del Título II se simplifica cuando se enuncian los objetivos del programa en términos de indicadores objetivamente verificables con metas de desempeño. Por ejemplo, un programa del Título II realizado en Mozambique en el período 1997-2001 requiere una mejora del almacenamiento en la finca y de la elaboración de alimentos en el año 2001 y especifica que eso se medirá en parte determinando si ha habido un aumento de 30% en la adopción de técnicas mejoradas de almacenamiento al final del proyecto.

3

Determinación de los requisitos del tamaño de la muestra

1. Factores que influyen en las decisiones sobre el tamaño de la muestra

El tamaño de la muestra necesaria para una encuesta dada se determina con sus objetivos de medición. En encuestas destinadas a medir ya sea los cambios de los indicadores con el tiempo o las diferencias de los indicadores entre las zonas del proyecto y las testigos, el tamaño de la muestra necesario para un indicador dado para cada ronda de la encuesta o grupo de comparación depende de cinco factores. Los dos primeros son las características demográficas y los tres últimos son escogidos por el evaluador o el diseñador de la encuesta. Son los siguientes:

- el número de unidades de medición en la población destinataria,
- el nivel inicial o *básico* del indicador,
- la magnitud del cambio o las diferencias entre los grupos de comparación que se espera medir con seguridad,
- el grado de confianza deseado para asegurarse de que cualquier cambio de la magnitud especificada antes o cualquier diferencia observada entre los grupos de comparación en ese sentido no habría ocurrido por casualidad (grado de significación estadística) y
- el grado de confianza deseado para asegurarse de detectar un cambio real o una diferencia de la magnitud especificada antes (poder estadístico).

Para ilustrar esa explicación, en el caso de una evaluación destinada a medir los cambios con el tiempo (por ejemplo, un diseño de una evaluación con una prueba *a priori* y otra *a posteriori* en un solo grupo), supongamos que se desea medir una reducción de 20 puntos porcentuales en la proporción de niños de 6 a 59 meses con insuficiencia de peso, con 95% de confianza y 80% de poder estadístico. Si una proporción estimada de 40% de los niños tuvieran insuficiencia de peso en el momento de la encuesta básica, el objetivo sería medir un cambio de la prevalencia de niños con insuficiencia de peso de 40% a 20% y tener (1) una confianza de 95% de que esa reducción no habría ocurrido por casualidad y (2) una confianza de 80% de detectar esa reducción si en realidad ocurrió (poder estadístico). Los cálculos del tamaño de la muestra permitirían responder a dos preguntas: (1) ¿cuántos niños de 6 a 59 meses (la unidad de medición) se necesitarían para lograr los objetivos enunciados? y (2) ¿cuántas familias habría que escoger para encontrar ese número de niños?

Para una evaluación con una prueba *a posteriori* solamente para comparar las zonas del proyecto y las testigos, supongamos que se desea medir la misma diferencia de 20 puntos porcentuales entre las dos con respecto a un indicador especificado. El tamaño de la muestra se fijaría de tal forma que permitiera detectar con fiabilidad esa diferencia entre las dos zonas. Se aplicaría el mismo principio en el caso de un diseño de una evaluación con una prueba *a priori* y otra *a posteriori* con zonas de tratamiento y testigos, con la excepción de que el tamaño de la muestra se fijaría de tal forma que permitiera la detección sistemática fiable de la diferencia en el grado de cambio.

2. Necesidades iniciales de información

Antes de iniciar el trabajo para determinar el tamaño de la muestra, hay que recoger información sobre dos asuntos:

- la composición de la familia y
- los niveles o tasas *esperados* o normales en los indicadores que se deben medir.

La composición de la familia se refiere a la proporción del total de familias que podrían tener una o varias personas en el subgrupo de interés. (El uso de esta información se explica más adelante en la sección 3.1.2, que describe la forma de convertir el número de elementos necesarios en el número de familias con las que se debe establecer contacto.) Por ejemplo, el indicador genérico de los programas del Título II “porcentaje de lactantes que reciben alimentos extra por dos semanas después de un episodio de diarrea” necesitará estimaciones de la proporción de familias con posibilidades de tener niños menores de 24 meses. La fuente habitual de información sobre la composición de la familia es el censo de población más reciente. En una situación ideal, se dispondrá de datos sobre la zona destinataria del programa evaluado. En caso contrario, se emplearían los datos del nivel de agregación siguiente (por ejemplo, distrito, provincia o región); si no se dispone de ellos, se pueden emplear los datos nacionales.

A menudo será difícil obtener información sobre los niveles o tasas esperados en los diversos indicadores que se pretende medir. Por ejemplo, al tratar de calcular el porcentaje de lactantes que recibieron alimentos extra después de un episodio de diarrea, se deben determinar dos cosas: (1) qué proporción de esos niños puede haber sufrido un episodio de diarrea en las dos semanas anteriores a la encuesta y (2) qué proporción puede haber recibido alimentos extra después de un episodio de diarrea durante el período justamente anterior a la encuesta. Las posibles fuentes de información son las encuestas anteriores que puedan haberse realizado en el país o en un país vecino, los datos del Ministerio de Salud o de otras entidades gubernamentales o los cálculos aproximados de personas informadas. Más adelante, en la sección 3.3, se ofrece orientación sobre lo que se debe hacer cuando no hay una fuente de información fidedigna.

La figura 3-1 ofrece ejemplos de la clase de datos preliminares sobre la población en cuestión de que se necesitan antes de poder iniciar el trabajo para establecer el tamaño de la muestra para una encuesta dada, con indicadores de alimentación del lactante y del niño pequeño en programas del Título II.

Figura 3-1: Ejemplo ilustrativo de las necesidades de información para determinar el tamaño de la muestra, con indicadores genéricos de alimentación del lactante y del niño pequeño en programas del Título II

A. Información sobre la composición demográfica:

1. Media de personas por familia
2. Proporción de la población total constituida por:
 - a. Niños menores de 0 a 59 meses.
 - b. Niños menores 24 meses.
 - c. Lactantes menores 6 meses.
 - d. Lactantes de 6 a 10 meses.

B. Información sobre los niveles o tasas *esperados* en la población destinataria:

1. Proporción de niños de 6 a 59 meses con retraso del crecimiento.
2. Proporción de niños de 6 a 59 meses con insuficiencia de peso.
3. Proporción de lactantes menores de 6 meses amamantados entre 1 y 8 horas después de nacer.
4. Proporción de lactantes menores de 6 meses alimentados solo con leche materna.
5. Proporción de lactantes de 6 a 10 meses que reciben alimentos complementarios.
6. Proporción de niños menores de 24 meses que tuvieron un episodio de diarrea en las 2 semanas anteriores a la encuesta.
7. Proporción de niños menores de 24 meses que tuvieron un episodio de diarrea en las 2 semanas anteriores a la encuesta y fueron alimentados continuamente.
8. Proporción de niños menores de 24 meses que tuvieron un episodio de diarrea en las 2 semanas anteriores a la encuesta y recibieron alimentos extra.

3. Cómputos del tamaño de la muestra

3.1 *Cálculo del número de elementos de muestra y familias*

Hay dos pasos para determinar los requisitos del tamaño de la muestra para una encuesta dada:

- 1) el cálculo del número de elementos de muestra necesarios para satisfacer los requisitos de medición con respecto a un indicador dado y
- 2) el cálculo del número de familias con las que habría que establecer contacto con el fin de encontrar el número de elementos necesarios en el primer paso.

Las fórmulas para esos cálculos se presentan más adelante en las secciones 3.1.1 y 3.1.2.

3.1.1 *Cálculo del número de elementos de muestra*

Los indicadores pueden expresarse como una proporción, una media o un total. En el primer caso, un indicador puede expresarse como una proporción de la población, por ejemplo, como el porcentaje de lactantes hasta de 6 meses alimentados exclusivamente con leche materna o con retraso del crecimiento (véase la lista de indicadores genéricos de alimentación del lactante y del niño pequeño en programas del Título II en la figura 3-6). En el segundo, cuando se expresan como una media o un total, se pueden enunciar según la cantidad de un producto o un bien en particular o como el número total de personas, por ejemplo, la ingestión media diaria de calorías per cápita en una población dada.

3.1.1.a. *Indicadores expresados como proporción*

La siguiente fórmula (ecuación básica 1) puede emplearse para calcular el tamaño de la muestra necesario para los indicadores expresados como porcentaje o como proporción. Nótese que el tamaño de las muestras obtenidas corresponde *a cada ronda de la encuesta o a cada grupo de comparación*.

Ecuación básica 1: Proporción

$$n = D [(Z_{\alpha} + Z_{\beta})^2 * (P_1 (1 - P_1) + P_2 (1 - P_2)) / (P_2 - P_1)^2]$$

CLAVE:

- n = tamaño mínimo de la muestra necesario por cada ronda de la encuesta o grupo de comparación,
- D = efecto del diseño (que en las siguientes ecuaciones se supone que es el valor predeterminado de 2; véase la sección 3.4 más adelante),
- P₁ = el nivel estimado de un indicador medido como proporción en el momento de la primera encuesta o para la zona testigo,
- P₂ = el nivel *esperado* del indicador ya sea en alguna fecha futura o para la zona del proyecto, de tal manera que la cantidad (P₂ - P₁) sea el tamaño de la magnitud del cambio que se aspira a detectar,
- Z_α = la puntuación Z correspondiente al grado de confianza con que se aspira a concluir que un cambio observado del tamaño (P₂ - P₁) no habría ocurrido por casualidad (α, el grado de significación estadística) y
- Z_β = la puntuación Z correspondiente al grado de confianza con que se aspira a detectar con seguridad un cambio de tamaño (P₂ - P₁), si realmente ocurrió (β, el poder estadístico).

Z_α y Z_β tienen un valor “estándar” según la fiabilidad deseada. Se presentan a continuación en la figura 3-2. Nótese que cuanto mayor sea el porcentaje, más seguridad se tendrá de medir resultados precisos en el programa.

Figura 3-2: Valores de Z_α y Z_β

α (alfa)	Z _α	β (beta)	Z _β
0,90	1,282	0,80	0,840
0,95	1,645	0,90	1,282
0,975	1,960	0,95	1,645
	2,326	0,975	1,960

0,99	0,999	2,320
------	-------	-------

El uso de la fórmula para calcular los indicadores expresados como proporción se presenta a continuación en la figura 3-3. Se incorporan los parámetros estándar Z_α y Z_β tomados de la figura 3-2.

Figura 3-3: Ejemplo ilustrativo de los cálculos del tamaño de la muestra para indicadores expresados como proporción

Ejemplo 1: Supongamos que se debe medir un aumento de 10 puntos porcentuales en la proporción de familias que muestran buenos hábitos de lavado de las manos. Supongamos además que en el momento de la primera encuesta, se creía que alrededor de 50% de las familias tenían buenos hábitos de lavado de las manos. En este caso, $P_1 = 0,50$ y $P_2 = 0,60$. Con los parámetros estándar de un grado de significación del nivel de 95% (α) y un poder estadístico de 80% (β), se escogen los valores de $Z_\alpha = 1,645$ y $Z_\beta = 0,840$ de la figura 3-2. Al incorporar esos valores en la fórmula anterior se logra el siguiente resultado:

$$\begin{aligned} n &= 2 [(1,645 + 0,840)^2 * ((0,5)(0,5) + (0,6)(0,4))] / (0,6 - 0,5)^2 \\ &= 2 [(6,175 * 0,49) / 0,10^2] \\ &= 2 [(3,02575) / 0,01] = 2 (302,575) = 605,15, \end{aligned}$$

o sea 606 familias por cada ronda de la encuesta.

Ejemplo 2: Supongamos que un programa de alimentación infantil espera aumentar 20 puntos porcentuales la proporción de lactantes menores de 6 meses que reciben solo leche materna en un período de cinco años. Supongamos además que, al principio, se cree que alrededor de 60% de los lactantes de la población destinataria son alimentados exclusivamente con leche materna durante seis meses. Por lo tanto, $P_1 = 0,60$ y $P_2 = 0,80$. Puesto que en el programa se desea detectar con bastante seguridad un aumento de 20 puntos porcentuales si realmente ocurriera, se escoge un poder estadístico de 90%, junto con el nivel normal de significación de 95%. De conformidad con ello, $Z_\alpha = 1,645$ y $Z_\beta = 1,282$. Por consiguiente, el tamaño de la muestra necesario es el siguiente:

$$\begin{aligned} n &= 2 [(1,645 + 1,282)^2 * ((0,6)(0,4) + (0,8)(0,2))] / (0,8 - 0,6)^2 \\ &= 2 [(8,567 * 0,40) / 0,20^2] \\ &= 2 [(3,4268) / 0,04] = 2 (86,67) = 171,34, \end{aligned}$$

o sea 172 lactantes per cada ronda de la encuesta.

La figura 3-4 permite que el diseñador de la encuesta que conoce la magnitud del cambio y el grado de precisión deseado escoja el tamaño de la muestra sin tener que realizar los cálculos precedentes. El tamaño de la muestra indicado se calculó empleando la fórmula básica en esta sección: los valores de los niveles iniciales del indicador (P_1) varían de 0,10 a 0,50 y los cambios y las diferencias en un indicador dado de magnitudes especificadas ($P_2 - P_1$) varían de 0,05 a 0,30. El cuadro es para valores de $\alpha = 0,95$ y $\beta = 0,80$. En el apéndice 2 se encuentra un cuadro comparable para $\alpha = 0,95$ y $\beta = 0,90$. En la sección 3.5 se ofrece orientación sobre la forma de escoger estos parámetros.

Figura 3-4: Tamaño de la muestra necesario para combinaciones selectas de P_1 y cambios o diferencias por detectarse entre los grupos de comparación ($\alpha = 0,95$ y $\beta = 0,80$)

Cambio/diferencia por detectarse ($P_2 - P_1$)												
P_1	Aumento						Reducción					
	0,05	0,10	0,15	0,20	0,25	0,30	0,05	0,10	0,15	0,20	0,25	0,30
0,10	1.074	309	152	93	63	45	679	111	21	\$	\$	\$
0,15	1.420	389	185	110	73	51	1.074	216	70	23	3	\$
0,20	1.717	457	213	124	81	56	1.420	309	114	49	21	7
0,25	1.964	513	235	134	86	60	1.717	389	152	73	37	19
0,30	2.161	556	251	142	90	62	1.964	457	185	93	51	29
0,35	2.310	587	262	147	92	62	2.161	513	213	110	63	38
0,40	2.408	605	268	148	92	62	2.310	556	235	124	73	45
0,45	2.458	611	268	147	90	60	2.408	587	251	134	81	51
0,50	2.458	605	262	142	86	56	2.458	605	262	142	86	56
0,55	2.408	587	251	134	81	51	2.458	611	268	147	90	60
0,60	2.310	556	235	124	73	45	2.408	605	268	148	92	62
0,65	2.161	513	213	110	63	38	2.310	587	262	147	92	62
0,70	1.964	457	185	93	51	29	2.161	556	251	142	90	62

0,75	1.717	389	152	73	37	19	1.964	513	235	134	86	60
0,80	1.420	309	114	49	21	7	1.717	457	213	124	81	56
0,85	1.074	216	70	23	3	\$	1.420	389	185	110	73	51
0,90	679	111	21	\$	\$	\$	1.074	309	152	93	63	45

3.1.1.b Para indicadores expresados como media o total

La siguiente fórmula puede emplearse para calcular los requisitos del tamaño de la muestra para indicadores expresados como media o total. Es preciso calcular el tamaño de la muestra para cada ronda de la encuesta o grupo de comparación.

Ecuación básica 2: Media o total

$$n = D [(Z_{\alpha} + Z_{\beta})^2 * (sd_1^2 + sd_2^2) / (X_2 - X_1)^2]$$

CLAVE:

- n = tamaño mínimo de la muestra necesario por cada ronda de la encuesta o grupo de comparación
- D = efecto del diseño para encuestas de conglomerados (útese el valor predeterminado de 2, como se discute en la sección 3.4)
- X₁ = el nivel estimado de un indicador en el momento de la primera encuesta o para la zona testigo
- X₂ = el nivel *esperado* del indicador ya sea en alguna fecha futura o para la zona del proyecto, de tal manera que la cantidad (X₂ - X₁) sea el tamaño de la magnitud del cambio o de las diferencias entre los grupos de comparación que se aspira a detectar
- sd₁ y sd₂ = desviaciones estándar *esperadas* de los indicadores en las respectivas rondas de la encuesta o en los grupos de comparación
- Z_α = la puntuación Z correspondiente al grado de confianza con que se aspira a concluir que un cambio observado del tamaño (X₂ - X₁) no habría ocurrido por casualidad (α, el grado de significación estadística) y
- Z_β = la puntuación Z correspondiente al grado de confianza con que se aspira a detectar con seguridad un cambio de tamaño (X₂ - X₁), si realmente ocurrió (β, el poder estadístico)

La principal dificultad del uso de la fórmula precedente está en que requiere información sobre la desviación estándar del indicador empleado en los cálculos del tamaño de la muestra. La solución preferida está en usar los valores de una encuesta anterior efectuada en el medio en que se realiza un programa objeto de evaluación. Si no se dispone de esos datos, se pueden emplear los de otra parte del país o los de un país vecino con características similares. Esos datos se presentan a menudo en los informes de las encuestas.

La fórmula precedente se aplica en la figura 3-5. Se emplean los valores estándar de Z_α y Z_β proporcionados en la figura 3-2.

Figura 3-5: Ejemplo ilustrativo del cálculo del tamaño de la muestra para indicadores expresados como media

Supongamos que se pretende medir los efectos de un programa del Título II en el consumo diario de calorías per cápita. Al comienzo del programa, se supone que la población destinataria del proyecto consume una media de 1.700 calorías diarias per cápita. La meta del programa es un aumento de 20%, o sea 340 calorías diarias per cápita. Por consiguiente, $X_1 = 1.700$ y $X_2 = 2.040$. Los datos de una encuesta reciente de una región vecina indican un consumo medio de 1.892 calorías diarias per cápita y una desviación estándar de 1.136.

Al calcular los requisitos del tamaño de la muestra, se supone que hay una relación constante entre la desviación estándar y la ingestión media de calorías en las dos rondas de la encuesta. Con los datos del país vecino, podemos hacer un cálculo aproximado de la desviación estándar empleando la relación entre la media del país vecino y la desviación estándar y aplicándola a las medias estimadas de X_1 y X_2 (1.700 y 2.040). Por lo tanto, la desviación estándar estimada (sd_1) se calcula estimando la relación entre la media de referencia o de la población vecina y la desviación estándar (1.892/1.136) o sea 1,6655 y aplicándola a la media estimada de X_1 : $1.700/1,6655 = 1.021$. La segunda desviación estándar (sd_2) se calcula aplicando la misma relación entre la media del país vecino y la desviación estándar (1.892/1.136) o sea 1,6655 y aplicándola a la media estimada de X_2 : $2.040/1,6655 = 1.225$.

A partir de los parámetros normales de un grado de significación de 95% y un poder estadístico de 80%, se escogen los valores de la figura 3-2 de $Z_\alpha = 1,645$ y $Z_\beta = 0,840$. Al incorporar esos valores a la fórmula precedente se obtiene el siguiente resultado:

$$n = 2[(1,645 + 0,840)^2 * (1021^2 + 1225^2) / (2.040 - 1.700)^2]$$

$$= 2[(6,175)(2.543.066) / (340)^2]$$

$$= 2[15.705.432 / 115.600] = 2(135,843) = 272$$

o sea 272 familias por cada ronda de la encuesta.

3.1.2 Determinación del número de familias con quienes se necesita establecer contacto

Los cálculos anteriores permiten que el diseñador de la encuesta averigüe con cuántos *elementos* de muestra necesita comunicarse para medir los cambios y las diferencias en los indicadores clave. Puesto que no todas las familias tendrán un miembro que se ajuste a la categoría señalada en el indicador (por ejemplo, niños menores de 6 meses o niños que han tenido algún episodio de diarrea en las dos últimas semanas), durante el trabajo de encuesta sobre el terreno normalmente se necesitará entrar en contacto con más familias de lo que indica el número de elementos de muestra. Entonces, el paso siguiente consiste en convertir los requisitos del tamaño de la muestra expresados como *elementos* en otros expresados como *familias*. (Nótese que no se necesita el procedimiento de cálculo del tamaño de la muestra en dos etapas para los indicadores medidos en cada familia, ya que, por definición, es posible medir esos indicadores en cada familia escogida en una muestra dada).

Por ejemplo, para el indicador “porcentaje de lactantes/niños <24 meses amamantados dentro de la primera hora de vida”, normalmente solo de 6 a 8% de la población de la zona destinataria del proyecto estará formada por lactantes/niños menores de 24 meses. (La escala indica el grado de fecundidad de la población; cuanto mayor sea la probabilidad, más alta será la proporción de lactantes y niños en la población). Por ende, se espera que en una población donde el promedio de personas por familia es de 6,0, de 36 a 48% de las familias tengan lactantes/niños menores de 24 meses.

El paso siguiente consiste en convertir esa información en el número de familias con las que se debe establecer contacto con el fin de encontrar el número de elementos necesario. Supongamos que en este caso se hubiera determinado que se necesitarían 300 lactantes/niños menores de 24 meses para producir el indicador de lactancia materna empleado en los ejemplos anteriores. Suponiendo además que esta es una población sumamente fecunda, entonces el número de familias se calcularía como $300 / (0,08 * 6)$ ó $300 / 0,48$, equivalente a un total $n = 625$ familias.

Nótese también que en la discusión precedente, se supone que *todos* los sujetos que cumplen los requisitos para determinado indicador encontrado en cada familia de la muestra deben incluirse en ésta (por ejemplo, todos los niños de 6 a 59 meses). Otra estrategia sería incluir en la muestra solamente un sujeto apto por familia (independientemente del número encontrado). Sin embargo, esto implicaría la necesidad de comunicarse con más familias y aumentaría los costos del trabajo de encuesta sobre el terreno. La estrategia de *inclusión de todos* también simplificará el cálculo de los valores de ponderación del muestreo, ya que no habrá que calcular la fracción de muestreo dentro de la familia (véase la fórmula de este cálculo en la figura 5-1, D). La desventaja de seleccionar a todos los sujetos que reúnan los requisitos dentro de las familias de muestra está en que a menos que se usen métodos más complejos de estimación de errores de muestreo, el grado estimado de error de muestreo en las estimaciones de la encuesta puede tener un sesgo hacia abajo, debido a la formación de conglomerados dentro de la familia. Como cuestión práctica, la magnitud de este sesgo suele ser pequeña y, a menos

que haya un estadígrafo y un programa de informática apropiado durante la etapa de análisis, este es un riesgo frecuentemente aceptado cuando se realiza una encuesta (véase el capítulo 5, sección 3).

No obstante, a veces, será necesario escoger solamente un sujeto por familia porque el muestreo de todos los sujetos que reúnan los requisitos prolongará demasiado las entrevistas de la encuesta. Por desgracia, raras veces se dispondrá de datos detallados sobre la proporción de familias que tienen sujetos que cumplen con los criterios establecidos para los diversos indicadores (si se llega a disponer de ellos algún día), lo que dificulta la determinación anticipada del número exacto de familias con las que habría que establecer contacto para lograr el tamaño de la muestra escogido como objetivo. Quizá lo mejor sea calcular el número de familias con las que se debe establecer contacto como se explicó antes y luego agregar una *reserva* de 20 a 25% de las familias a manera de compensación porque, en algunos casos, se dejará de escoger a algunos de los sujetos que reúnen los requisitos dentro de una familia. No obstante, es preciso reconocer que éste es solamente un cálculo aproximado.

La figura 3-6 ofrece orientación sobre el número típico de familias con que se deberá establecer contacto para obtener una sola unidad de medición para cada indicador genérico de salud dentro de los programas del Título II.

Figura 3-6: Número típico de familias por entrevistar en búsqueda de una persona de referencia para los indicadores genéricos de salud en programas del Título II (con la suposición de que hay seis personas por familia)			
Indicador	Persona de referencia	Porcentaje de la población ¹	No. necesario de familias ¹
% con retraso del crecimiento	niños de 6 a 59 meses	14-19%	1,2/0,9
% con insuficiencia de peso	niños de 6 a 59 meses	14-19%	1,2/0,9
% amamantados en las 8 horas siguientes al nacimiento	lactantes < 24 meses	6-8%	2,8/2,1
% alimentados exclusivamente con leche materna	lactantes < 6 meses	1,5-2,0%	11,1/8,3
% que recibieron	lactantes de 6 a 10	1,0-1,5%	16,7/11,1

alimentos complementarios	meses		
% con episodio de diarrea en las 2 últimas semanas	niños < 24 meses	6-8%	2,8-2,1
% con episodio de diarrea que recibieron alimentación continua	niños < 24 meses con episodio de diarrea	1,5-2,0% ²	11,1/8,3
% con episodio de diarrea que recibieron alimentación extra	niños < 24 meses con episodio de diarrea	1,5-2,0% ²	11,1/8,3
<p>1. Las escalas presentadas son para poblaciones con bajas y altas tasas de fecundidad, respectivamente.</p> <p>2. Se supone que hay una frecuencia de diarrea de 25% en las dos semanas precedentes. Se pueden emplear estimaciones locales como sustituto.</p>			

3.2 Selección de indicadores para determinar los requisitos del tamaño de la muestra

En una encuesta típica para una evaluación de un programa del Título II se medirán varios indicadores. En condiciones ideales, se considerarían los requisitos establecidos para cada indicador al determinar el tamaño de la muestra necesario para una encuesta dada. No obstante, cuando el número de indicadores que debe medirse es grande, esta sería una operación engorrosa.

Este problema suele abordarse de una de dos maneras. Una opción consiste en determinar cuál de los indicadores puede ser el más exigente en lo que respecta al tamaño de la muestra y emplear el tamaño de la muestra que exige ese indicador. Al hacerlo, se cumplirá con los requisitos establecidos para los demás indicadores. En la mayoría de los casos, ese será el indicador cuya unidad de medición es la encontrada con menos frecuencia en la población destinataria. Por ejemplo, en la figura 3-6, ese será el “porcentaje de lactantes de 6 a 10 meses que recibieron alimentos complementarios”, ya que solamente alrededor de 1,0 a 1,5% de la población destinataria estará constituida por lactantes de ese grupo de edad. La principal ventaja de este procedimiento radica en que asegurará automáticamente el uso de un tamaño adecuado de la muestra para todos los indicadores que se deben medir. El inconveniente está en que se escogerá un tamaño de la muestra mayor de lo necesario para algunos o para muchos indicadores.

Un segundo enfoque consistiría en buscar un pequeño número de indicadores que se consideren de mayor importancia para fines de evaluación del programa y limitar los cálculos del tamaño de la muestra a esos indicadores. Eso permitirá asegurar el uso de un tamaño adecuado de la muestra para los indicadores clave. El inconveniente está en la posibilidad de dejar de obtener un tamaño adecuado de la muestra para otros indicadores que pueden ser más exigentes en lo que respecta a los requisitos del tamaño de la muestra. Sin embargo, como dichos requisitos ya se han considerado de importancia secundaria, este puede ser un compromiso razonable, especialmente cuando hay limitaciones de tiempo y de recursos.

Otro compromiso, cuando hay que tener en cuenta varias cuestiones financieras y logísticas, consiste en calcular los requisitos del tamaño de la muestra para los indicadores clave y para el indicador más exigente en lo que respecta al tamaño de la muestra y luego escoger entre los dos el tamaño más grande posible. Esta opción garantiza un tamaño adecuado de la muestra para los indicadores clave y las mejores estimaciones posibles para los indicadores más exigentes, dados los recursos disponibles. Este enfoque se usa a menudo.

3.3 Selección de los valores iniciales desconocidos de los indicadores

En una situación ideal, el valor *básico* de un indicador expresado como proporción, es decir, P_1 , se obtiene a partir de la información disponible de otras encuestas realizadas en un medio determinado (por ejemplo, información de encuestas anteriores que pueden haberse realizado en la región o en un país vecino; datos del Ministerio de Salud o de otras entidades gubernamentales; o, cuando no se dispone de esa información, cálculos aproximados de personas informadas, con base en las mejores fuentes disponibles, como se indicó en la sección 2 precedente). Al escoger un valor para P_1 , es mejor inclinarse hacia un valor de 0,50. La razón de ello está en que la varianza de los indicadores medidos como proporción alcanza su punto máximo a medida que se acercan a 0,50. Lo más seguro sería escoger siempre $P_1 = 0,5$, ya que ello asegurará un tamaño adecuado de la muestra, independientemente del valor real de P_1 . No obstante, eso dará como resultado muestras más grandes de lo necesario en caso de que el valor real de P_1 sea muy diferente de 0,50. Por ende, se recomienda hacer la mejor estimación a partir de la información disponible e inclinarse hacia la selección del valor de P_1 más cercano a 0,50. Por ejemplo, si se pensara que un indicador está en la escala de 0,30 a 0,40 en el momento de la encuesta de base, se escogería 0,40.

Para los indicadores expresados como media o total, se necesitan estimaciones no solamente para obtener el valor básico en el momento de la primera encuesta o en la zona testigo (X_1), sino también para la desviación estándar de X . En este caso el parámetro de importancia decisiva es la desviación estándar. Por desgracia, no hay una regla empírica simple en caso de que no se disponga de datos de otras encuestas. Se aconseja ser conservador y escoger un valor de la desviación estándar que sea una proporción considerable del valor inicial presunto del indicador; por ejemplo, las desviaciones estándar

que representen de 60 a 80% de la estimación de trabajo de X deben ser adecuadas en la mayoría de los casos.

3.4 Efectos del diseño

En las dos ecuaciones básicas precedentes, “D” denota el efecto del diseño. Esto ofrece una forma de corregir la pérdida de eficiencia del muestreo que se produce al usar el muestreo por conglomerados en lugar del muestreo aleatorio simple (véase el capítulo 4). Puede considerarse como el factor por el cual habría que aumentar el tamaño de una muestra por conglomerados para producir estimaciones en la encuesta con la misma precisión que con una muestra aleatoria simple. La magnitud de D depende de dos factores: (1) el grado de similitud o de homogeneidad de los elementos dentro de los conglomerados y (2) el número de unidades de medición que se tomarán de cada conglomerado.

En una situación ideal, podría obtenerse una estimación de D para los indicadores de interés a partir de una encuesta anterior en un medio determinado. Eso dará una idea de la similitud u homogeneidad de los elementos del conglomerado. A falta de ello, se podrían usar los valores *típicos* de las encuestas realizadas en otra parte. Lamentablemente, esa orientación falta a menudo y, por ende, suele emplearse un valor predeterminado de 2,0, especialmente para las encuestas antropométricas y de inmunización. Suponiendo que el tamaño de la muestra del conglomerado se pueda mantener moderadamente pequeño (véase una discusión más detallada sobre los diversos tamaños del conglomerado en la sección 4 más adelante), el uso de un valor estándar de $D = 2,0$ debe compensar suficientemente el uso del muestreo por conglomerados en la mayoría de los casos.

3.5 Significación y poder estadísticos

Se puede pensar que la significación (α) y el poder (β) estadísticos son análogos de positivos falsos y negativos falsos: la significación estadística protege contra la falsa conclusión de que ha ocurrido un cambio, en tanto que el poder estadístico protege contra la falsa conclusión de que no ha sucedido nada como resultado de un programa. De los dos, quizá el más importante para las evaluaciones de los programas sea el parámetro del poder estadístico, β , ya que asegura que no se juzgue un programa como fracaso cuando, de hecho, ha tenido un resultado positivo. A menos que el tamaño de la muestra sea suficiente para poder detectar con fiabilidad los cambios o las diferencias entre los grupos de comparación de un tamaño especificado, queda comprometida la utilidad de las encuestas como instrumento de evaluación de programas. Cuando el poder estadístico no es suficiente puede llevar a una falsa conclusión de que no hubo cambios significativos en los indicadores con el tiempo ni diferencias entre los grupos del proyecto y testigos, cuando, de hecho, hubo cambios y diferencias *reales* indetectables porque el tamaño de la muestra usada no fue suficiente. Para asegurarse de tener suficiente poder estadístico, es preciso usar un valor mínimo de β de 0,80 y es preferible usar 0,90 donde lo permitan los recursos.

En el caso de α , o sea el grado de significación, el valor normal en la mayoría de las encuestas es de 95%; se supone que esto es suficiente para asegurarse de que cualquier cambio observado no haya ocurrido por casualidad. Sin embargo, si los recursos no lo permiten, este parámetro podría reducirse a $\alpha = 0,90$. Esta cifra menor entraña solamente un modesto riesgo adicional de llegar a la falsa conclusión de que ha ocurrido un cambio o de que los indicadores de los grupos del proyecto y testigos son diferentes. Sin embargo, no se recomiendan valores inferiores a este grado de significación y en figura 3-2 no se proporcionan equivalencias de una cifra menor.

3.6 Margen para la falta de respuesta

La falta de respuesta es un hecho reconocido en las encuestas. Aunque se recomienda encarecidamente hacer todo lo posible por reducir al mínimo la falta de respuesta (véase el capítulo 4), hay límites prácticos sobre lo que se puede hacer.

Para poder lograr el tamaño de la muestra escogido como objetivo para una encuesta, se suele dejar un margen para la falta de respuesta en el cálculo de los requisitos del tamaño de la muestra. Esto implica normalmente el aumento del tamaño de la muestra por un factor que represente un *seguro* contra la falta de respuesta. Aunque eso variará algo de un medio a otro, un margen de 10% debe ser suficiente en la mayoría de las situaciones. Por ende, si el tamaño de la muestra calculado para una encuesta exigiera $n = 1.000$ familias y se incorporara una reserva de 10% para la falta de respuesta en el diseño de la muestra, el tamaño corregido de la muestra escogida como objetivo para la encuesta sería de $n = 1.100$ familias.

4. Determinación del número de conglomerados y del número de sujetos por conglomerado que se deben escoger

Una vez determinados los requisitos generales del tamaño de la muestra, el paso final para la creación del diseño de la muestra consiste en determinar cuántos conglomerados y cuántas familias por conglomerado deben escogerse. Esto exige tres consideraciones principales:

- La primera es la magnitud del efecto del diseño (D) del muestreo por conglomerados. Cuanto menor sea el número de familias por conglomerado, menos pronunciado será el efecto del diseño. Eso se debe a que, por lo general, las unidades elementales dentro de los conglomerados tienden a presentar un cierto grado de homogeneidad en lo que respecta a las características básicas y quizá a los patrones de comportamiento. A medida que aumenta el número de familias por conglomerados, se pierde la precisión del muestreo.

- En segundo lugar, el número de familias en un cierto conglomerado o sitio impone un límite sobre el tamaño potencial de la muestra por conglomerado. Las listas del censo u otros materiales que se pretenda emplear como marco de muestreo deben revisarse con cuidado antes de decidir sobre el tamaño de la muestra por conglomerado que se pretenda usar.
- Tercero, los recursos disponibles para realizar el trabajo de encuesta sobre el terreno determinan lo que es factible. El transporte y mantenimiento del personal y de los supervisores asignados al terreno constituyen los principales costos de la realización del trabajo de encuesta allá y tienden a variar más o menos directamente con el número de conglomerados que debe cubrirse. De conformidad con ello, los costos del trabajo sobre el terreno se reducen al mínimo cuando se mantiene pequeño el número de conglomerados.

Puesto que las dos últimas consideraciones pueden variar mucho en diferentes aplicaciones y medios, aquí solo puede ofrecerse orientación general. Desde el punto de vista de la precisión del muestreo, se prefieren los conglomerados pequeños a los grandes. Por consiguiente, cuando el tamaño de la muestra escogido como objetivo es fijo (por ejemplo, 600 familias), se preferiría un diseño con 30 conglomerados de 20 familias cada uno a otro con 20 conglomerados de 30 familias cada uno que, a su vez, se preferiría a otro con 10 conglomerados de 60 familias cada uno. Por regla general, la selección de un máximo de 40 a 50 familias por conglomerado sería relativamente segura. Por supuesto, si los recursos no permiten tener conglomerados de ese tamaño, se podría aumentar el número de unidades de medición que se toma en cada conglomerado, pero cabe reconocer que esa medida se tomará a costa de un mayor error de muestreo.

Aunque el uso de 30 conglomerados en encuestas basadas en la población se ha popularizado, de hecho, no hay justificación estadística de que 30 sea el número mínimo o ideal. Sin embargo, sirve de directriz aproximada de trabajo y representa una cifra adecuada para asegurarse de que las muestras de los miembros del grupo escogido como objetivo estén suficientemente dispersas en un número suficiente de conglomerados, de manera que en las estimaciones de la encuesta no influya indebidamente un puñado de conglomerados.

Empleando un número de 50 familias como el tamaño estándar de la muestra por conglomerado, el número de conglomerados se calcula normalmente dividiendo el tamaño de la muestra por 50. Por ejemplo, si se necesitara un tamaño de la muestra de $n = 2.000$ familias, la muestra se dispersaría en 40 conglomerados de 50 familias cada uno. Si el tamaño de la muestra es demasiado pequeño ($n = 1.000$, por ejemplo), habrá demasiado pocos conglomerados (en este caso, 20) y el tamaño de la muestra no se dispersará lo suficiente. En esa situación convendría tomar 34 conglomerados de 30 familias (aunque con eso se exceda un poco el tamaño de la muestra escogido como objetivo).

Es mejor que cada conglomerado tenga el mismo número de elementos de muestra. En primer lugar, eso asegura más o menos la misma carga de trabajo en cada conglomerado, lo que facilita hasta cierto punto el control operativo del trabajo de la encuesta realizado sobre el terreno. En segundo lugar, se

evita el sesgo en la estimación ayudando a que la muestra tenga un mecanismo de autoponderación, tema discutido en detalle en los capítulos 4 y 5.

5. Requisitos del tamaño de la muestra para las encuestas de seguimiento

Los procedimientos aquí explicados para determinar el tamaño de la muestra para las encuestas se han establecido con el fin de tener en cuenta los requisitos para una ronda de encuestas de seguimiento como parte de la evaluación de un programa. No obstante, en algunos casos, deberá aumentarse el tamaño de la muestra en la segunda ronda. En algunos casos, ocurrirá eso cuando, al determinar el tamaño de la muestra para la primera ronda de la encuesta, no se tuvieron en cuenta los requisitos para medir el cambio con el tiempo; por ejemplo, si la fórmula empleada para calcular el tamaño de la muestra para una encuesta única se usó para determinar el tamaño de la muestra para la encuesta básica. En otros, ocurrirá cuando se usó una fórmula apropiada para calcular el tamaño de la muestra, pero los niveles de los indicadores observados en la encuesta básica fueron diferentes de los esperados al hacer los cálculos del tamaño de la muestra antes de la encuesta. Específicamente, en casos en que el indicador se midió como proporción, el valor observado en la encuesta básica resultó ser mucho más cercano a 0,50 de lo esperado; cuando se midió el indicador como media o total, la desviación estándar del indicador resultó ser mucho mayor de lo previsto. En ambos casos, el tamaño de la muestra empleado para la encuesta básica sería demasiado pequeño para cumplir con los requisitos de precisión para la actividad de evaluación al emplearlo para la encuesta de seguimiento.

Hay dos maneras de calcular cuánto se necesita aumentar la muestra para la encuesta de seguimiento. La primera consiste en usar una de las fórmulas existentes, pero eso es problemático porque en éstas se tiene en cuenta la significación estadística, pero no el poder estadístico. Más bien, la solución recomendada es (a) computar una estimación revisada del requisito del tamaño de la muestra con la ecuación básica 1 ó 2, teniendo en cuenta los resultados de la encuesta básica y (b) compensar cualquier deficiencia del tamaño de la muestra en la encuesta básica con un aumento del tamaño de la muestra para la encuesta de seguimiento. Este procedimiento de dos pasos se presenta en las figuras 3-7 y 3-8.

Figura 3-7: Ejemplo ilustrativo del cálculo del tamaño de la muestra de la encuesta de seguimiento para indicadores expresados como proporción

Supongamos que se deseara medir un aumento de 10 puntos porcentuales en la proporción de familias con hábitos apropiados de lavado de las manos. En el momento de la encuesta básica se pensó que alrededor de 30% de las familias tenían hábitos apropiados de lavado de las manos. Por consiguiente, P_1 y P_2 se fijaron en 0,30 y 0,40, respectivamente. Con la fórmula para el cálculo de indicadores expresados como proporción (ecuación básica 1) y suponiendo que hay parámetros estándar de un grado de significación de 95% y un poder estadístico de 80%, se calcula un tamaño de la muestra de $n = 556$.

Sin embargo, la encuesta básica reveló que una proporción de familias mucho mayor (50%) ya tenía un hábito apropiado de lavado de las manos. El procedimiento aquí recomendado exige revisión del requisito del tamaño de la muestra con los resultados de la encuesta básica, es decir, $P_1 = 0,50$, de la manera siguiente:

$$\begin{aligned}
 n &= D [(Z_{\alpha} + Z_{\beta})^2 * (P_1 (1 - P_1) + P_2 (1 - P_2)) / (P_2 - P_1)^2] \\
 n &= 2 [(1,645 + 0,840)^2 * ((0,5)(0,5) + (0,6)(0,4))] / (0,6 - 0,5)^2 \\
 &= 2 [(6,175 * 0,49) / 0,10^2] \\
 &= 2 [(3,02575) / 0,01] = 2 (302,575) = 605,15 \text{ o sea } 606 \text{ familias por cada ronda de la encuesta.}
 \end{aligned}$$

Para compensar la deficiencia del tamaño de la muestra en la encuesta básica, se agrega la diferencia entre las estimaciones iniciales y revisadas del tamaño de la muestra (denotadas como n_1 y n_2 , respectivamente) al tamaño de la muestra que se pretende usar en la encuesta de seguimiento. Por lo tanto, el tamaño de la muestra para la encuesta de seguimiento sería:

$$n = 606 + (n_2 - n_1) = 606 + (606 - 556) = 606 + 50 = 656$$

Figura 3-8: Ejemplo ilustrativo del cálculo del tamaño de la muestra de la encuesta de seguimiento para indicadores expresados como media

En la figura 3-5, se hicieron los cálculos del tamaño de la muestra en encuestas básicas o de seguimiento para evaluar los efectos de un programa del Título II en el consumo diario de calorías per cápita. Supongamos que en lugar de un consumo medio diario esperado de calorías per cápita de $X_1 = 1.700$ y una desviación estándar de $sd_1 = 1.020$, se hubieran observado valores de $X_1 = 1.800$ y $sd_1 = 1.280$ en la encuesta básica. Siguiendo el procedimiento recomendado, se volverían a computar los requisitos del tamaño de la muestra de la manera siguiente (empleando parámetros estándar de un grado de significación de 95% y un poder estadístico de 80% y recalculando sd_2 con la proporción de X_1 / sd_1 observada en la encuesta básica):

$$\begin{aligned}n &= D [(Z_\alpha + Z_\beta)^2 * (sd_1^2 + sd_2^2) / (X_2 - X_1)^2] \\n &= 2[(1,645 + 0,840)^2 * (1.280^2 + 1.536^2) / (2.160 - 1.800)^2] \\&= 2[(6,175)(3.997.696) / (360)^2] \\&= 2[24.686.672 / 129.600] = 2(190.484) = 380,97,\end{aligned}$$

o sea 381 familias por cada ronda de la encuesta.

Para compensar la deficiencia del tamaño de la muestra en la encuesta básica, se agregaría luego la diferencia entre las estimaciones iniciales ($n = 272$) y revisadas del tamaño de la muestra (denotadas como n_1 y n_2 , respectivamente) al tamaño de la muestra que se pretende usar para la encuesta de seguimiento. Por consiguiente, el tamaño de la muestra escogido como objetivo para la encuesta de seguimiento sería:

$$n = 381 + (n_2 - n_1) = 381 + (381 - 272) = 381 + 109 = 490$$

4

Selección de la muestra

1. Perspectiva general del muestreo

1.1 Métodos de muestreo probabilístico y no probabilístico

Los procedimientos de muestreo son de dos clases: métodos formales o probabilísticos y métodos informales o no probabilísticos.

- **Los métodos formales de muestreo** se basan en la teoría del muestreo probabilístico. Éste exige dos cosas: (1) que cada unidad de muestreo tenga una probabilidad de selección conocida o sin valor de cero en la muestra y (2) que la posibilidad aleatoria sea el factor de control en la selección de unidades de muestreo. En la práctica, el muestreo probabilístico también tiende a caracterizarse por (1) el uso de listas o marcos de muestreo para seleccionar la muestra, (2) procedimientos claramente definidos de selección de la muestra y (3) la posibilidad de estimar los errores de muestreo a partir de los datos de la encuesta (discutidos con más detalle en el capítulo 5, sección 3).
- **Los métodos informales de muestreo** comprenden varios sistemas basados en principios distintos de los de probabilidad. Aunque la intención general que se tiene a menudo es hacer inferencias a alguna población mayor, los métodos de selección tienden a ser más subjetivos. En la mayoría de los casos, se supone que las personas que hacen la selección de la muestra están bien informadas sobre las dimensiones fundamentales en las que varían los fenómenos objeto de estudio y, por ende, pueden seleccionar la muestra de tal manera que éstos queden debidamente *cubiertos* (por ejemplo, que estén exentos de sesgo). Se pretende y se espera que la muestra sea suficientemente representativa para los fines de la encuesta, pero eso no se puede determinar con ningún grado mensurable de seguridad. El muestreo por *cuotas* y el muestreo *deliberado* constituyen dos ejemplos de las diversas formas de muestreo informal o no probabilístico (véase lo indicado más adelante en las secciones 3.2 y 3.5).

Se recomiendan enfáticamente los métodos de muestreo probabilístico para las evaluaciones de los programas del Título II, a pesar de que su costo es un poco alto. Por supuesto, se reconoce que ambos tipos de métodos de muestreo podrían producir los mismos resultados. Sin embargo, como se basan en la teoría estadística, las evaluaciones hechas con muestreo probabilístico tienen un mayor grado de credibilidad y se pueden defender más fácilmente que las basadas en los métodos informales de

muestreo, que pueden ser vulnerables a preguntas sobre si la muestra es representativa de la población o si está sesgada.

Todos los procedimientos de muestreo descritos en esta guía se basan en principios de probabilidad. Se han adaptado para tener en cuenta la clase de dificultades encontradas a menudo en el medio común de los países en desarrollo.

1.2 Muestreo por conglomerados

El muestreo por conglomerados tiende a ser el tipo de muestreo probabilístico de uso más amplio. De ordinario se escoge en lugar del muestreo aleatorio simple, aunque este es el método de muestreo probabilístico más directo y bien conocido. El muestreo aleatorio, que consiste en escoger las unidades individual y directamente por un proceso al azar en que cada unidad tiene la misma probabilidad de ser escogida, exige listas completas de unidades elementales (por ejemplo, familias en la zona del proyecto). Puesto que raras veces se dispone de ellas y, en general, el costo de crearlas es prohibitivo, es poco probable que el muestreo aleatorio se use como método “autónomo” de muestreo para la evaluación de programas del Título II. Por contraste, el muestreo por conglomerados limita el alcance de la construcción del marco de la muestra y el trabajo sobre el terreno a un subgrupo o a una muestra de regiones geográficas que se deben cubrir y, por ende, ofrece una forma de controlar los costos sobre el terreno.

Un **conglomerado** es sencillamente una agregación de unidades de muestreo de interés para una encuesta particular que se puede definir sin ambigüedad y emplear como unidad de muestreo de la cual se puede seleccionar una submuestra más pequeña. En condiciones ideales, los conglomerados deben ceñirse a cuatro criterios. (1) Deben tener límites físicos relativamente claros para facilitar la identificación sobre el terreno. (2) Deben estar localizados cerca unos de otros; de lo contrario, los costos subirán en exceso, lo que sería contrario a la principal finalidad del muestreo por conglomerados. (3) Los conglomerados no deben incluir a demasiadas personas; eso ayudará a reducir al mínimo el trabajo pendiente de creación del marco de muestreo. (4) En situaciones ideales, es preciso disponer de la información sobre el tamaño del conglomerado antes de seleccionarla. Eso permitirá usar los procedimientos de selección de la muestra destinados a mejorar la eficiencia del muestreo, o la selección con una *probabilidad proporcional al tamaño* (PPS) (véase la sección 3.1 más adelante). (La imposibilidad de obtener medidas del tamaño del conglomerado antes de la selección de la muestra no impide usar el método de muestreo probabilístico).

La creación del marco de muestreo entraña normalmente dos pasos: (1) selección de unidades de la primera etapa o *unidades primarias* y (2) selección de unidades elementales de muestreo dentro de las unidades primarias. Por ejemplo, en muchas aplicaciones de este método, se escogerán manzanas de poblados o de ciudades en la primera etapa y una muestra de familias de cada una en la segunda. En algunos casos, quizá sea necesario seleccionar a algunos miembros de las familias, con lo que se agrega un tercer paso al proceso.

Cuando se considere que la población de una unidad de muestra es demasiado grande, la selección de las unidades de muestreo de la *primera etapa* o *primarias* puede entrañar dos pasos. En esos casos, el conglomerado seleccionado se divide en dos o más conglomerados de menor tamaño, se selecciona al azar uno de los más pequeños y se realizan las operaciones de creación del marco de la muestra y de muestreo dentro del conglomerado más pequeño que se haya seleccionado. Puede haber un problema de sesgo en la selección del conglomerado más pequeño.

2. Marcos de muestreo

Un marco de muestreo es una lista de posibles unidades de muestreo. Para los tipos de encuestas considerados en este capítulo, por lo general, los marcos de muestreo son listas de zonas de enumeración del censo, correspondientes aproximadamente a manzanas de pueblos y ciudades, tomadas del último censo de población. Cuando no existen esas listas, bastará con una lista de pueblos y ciudades pequeñas que cubra a toda la población objetivo del proyecto, pero se pueden necesitar otras etapas de selección de la muestra para producir unidades suficientemente pequeñas con las que se pueda trabajar en las operaciones de la encuesta realizadas sobre el terreno.

Cuando no existe un marco adecuado de muestreo, hay dos alternativas. Una consiste en crear un marco inicial de muestra para la zona en cuestión. Por ejemplo, si no se dispone de una lista de pueblos para un proyecto que cubra a toda una provincia, se puede empezar por tomar una muestra de los distritos localizados dentro de la provincia y establecer un marco de muestreo de los pueblos solamente en los distritos de muestra. Este tipo de muestreo por conglomerados en varias etapas se emplea a menudo para compensar las deficiencias de los marcos de muestreo. En las encuestas es común crear marcos de muestreo de algún tipo o actualizar el trabajo y eso debe hacerse de tal forma que implique un gasto de tiempo y un costo mínimos.

Una segunda opción consistiría en restringir la encuesta a la parte del universo de encuesta para la cual existe el marco de muestreo. Sin embargo, esto limitaría la capacidad de generalizar los resultados de la evaluación a la parte de la zona del proyecto cubierta en el ejercicio de evaluación, que puede introducir un sesgo. Esta opción debe escogerse solamente como último recurso.

3. Procedimientos de selección de muestras para los programas con un alto grado de cobertura de la población en general

En algunos casos, se puede esperar razonablemente que una proporción considerable de la población en la zona del proyecto escogida como objetivo haya estado *expuesta* a la intervención objeto de evaluación (por ejemplo, que haya recibido o, por lo menos, tenido la oportunidad de recibir los beneficios del proyecto). Por ejemplo, los programas que influyen en las prácticas de alimentación del lactante se destinan comúnmente a la población en general de una región geográfica escogida para una

intervención. En esos casos, las encuestas de la población en general representan un medio válido de medir los efectos del programa.

A continuación se incluyen las indicaciones para seleccionar muestras en cada etapa del proceso de muestreo: (1) seleccionar los conglomerados de muestra, (2) seleccionar las familias de muestra y (3) seleccionar los sujetos particulares para la encuesta cuando se presente la necesidad. Se describen diferentes variantes para cada etapa.

3.1 Procedimiento de selección de los conglomerados de muestra

En el caso de los proyectos que ofrecen cobertura a la población en general, es posible que se enumeren demasiados conglomerados en el marco de muestreo para permitir un muestreo completo. Por lo tanto, como primer paso se puede reducir el número de conglomerados que se deben seleccionar a un número que permita realizar el trabajo. El procedimiento recomendado, **el muestreo aleatorio sistemático**, entraña escoger un conglomerado de una muestra al azar y cada i° conglomerado de la serie de ahí en adelante. (El i° se refiere al número de cada conglomerado escogido; por ejemplo, en la figura 4-2, se escogen los conglomerados primero, quinto y octavo). Eso puede hacerse de dos maneras, según se disponga de información sobre el tamaño de los conglomerados en la población destinataria. Ambos métodos se describen a continuación.

Cuando se dispone de las medidas del tamaño del conglomerado

El diseño del conglomerado de máxima eficiencia estadística en dos etapas es aquel en que (1) se seleccionan los conglomerados con probabilidad proporcional al tamaño (PPS) en la primera etapa de selección de la muestra y (2) se escoge un número constante de familias de cada conglomerado en la segunda etapa.

La expresión *probabilidad proporcional al tamaño* (PPS) significa que se da a los conglomerados más grandes una mayor posibilidad de selección que a los más pequeños. El uso del procedimiento de selección con PPS exige disponibilidad de un marco de muestreo por conglomerados con medidas del tamaño o su creación antes de seleccionar las muestras. Una *medida del tamaño* es sencillamente un recuento o una estimación de una variable que tiene posibilidades de correlacionarse con el número de los sujetos de la encuesta de interés en cierto conglomerado; por ejemplo, tanto la población total de un pueblo como el número total de familias puede tener una estrecha correlación con el número de lactantes/niños en el conglomerado y, por ende, son buenas medidas del tamaño para las encuestas en que los lactantes/niños serán importantes unidades de medición. No se necesita un recuento exacto; bastará con cálculos o estimaciones aproximados. Cualquier inexactitud se corregirá en la segunda etapa de selección de la muestra, cuando se escogerá el número específico de familias (véase la sección 3.2 a continuación).

En la figura 4-1 se enumeran los pasos para seleccionar una muestra de conglomerados con un muestreo sistemático con probabilidad proporcional al tamaño. En la figura 4-2 se aplican esos pasos a un ejemplo ilustrativo en que el objetivo es seleccionar 40 conglomerados de un total de 170.

Figura 4-1: Pasos para seleccionar una muestra aleatoria sistemática de conglomerados con PPS

- (1) Preparar una lista de unidades de muestreo para la primera etapa (es decir, de conglomerados) con la medida correspondiente del tamaño de cada una (véase la columna 2 en la figura 4-2).
- (2) Comenzando en la parte superior de la lista, calcular la medida acumulativa del tamaño y anotar esas cifras en una columna al lado de la medida del tamaño de cada unidad (véase la columna 3).
- (3) Calcular el intervalo de muestreo (SI) dividiendo la medida acumulativa total del tamaño correspondiente al dominio o estrato (M) por el número planeado de unidades que deben seleccionarse (a), es decir, $SI = M/a$.
- (4) Seleccionar un número al azar (comienzo al azar o RS) entre 1 y (SI). Comparar ese número con la columna de la medida acumulativa del tamaño. La unidad que tenga una medida acumulativa del tamaño dentro de la cual esté el número (RS) es la primera unidad de muestra (véase la columna 4).
- (5) Las unidades subsiguientes se escogen agregando el intervalo de muestreo (SI) al número identificado en el paso (4); es decir, $RS + SI$, $RS + SI * 2$, $RS + SI * 3$, etc. (véase la columna 4).
- (6) Se sigue este procedimiento hasta acabar con la lista. El número resultante de unidades debe ser aproximadamente igual al número de conglomerados escogido como objetivo.

Figura 4-2: Ejemplo ilustrativo: selección de una muestra aleatoria sistemática de conglomerados con PPS

Conglomerado No.	Tamaño - No. de familias	Medida acumulativa del tamaño	Muestreo No.	Conglomerado seleccionado
001	120	120	73	X
002	105	225		
003	132	357		
004	96	453		
005	110	563	503	X
006	102	665		
007	165	839		
008	98	937	934	X
009	115	1.052		
.	.	.		
.	.	.		
.	.	.		
170 (último)	196	17.219 (M en la ecuación)		
	Medida acumulativa total del tamaño	17.219		

No. planeado de conglomerados = 40 (a en la ecuación)
Intervalo de muestreo = $17.219/40 = 430.475$
Comienzo al azar entre 1 y 430.475 = 73
Conglomerados seleccionados = 001, 005, 008,

Siempre que sea posible, se deben seleccionar los conglomerados con probabilidad proporcional al tamaño en las encuestas de muestra. En primer lugar, porque este procedimiento es relativamente eficiente en lo que respecta a la precisión del muestreo. En segundo lugar, porque si se escoge un número igual de elementos en cada conglomerado en la segunda etapa de selección de la muestra, el resultado final será una muestra en que cada familia tiene la misma probabilidad general de selección, es decir, una muestra con un mecanismo de *autoponderación*. Esta es una gran ventaja durante el análisis de datos (véase una discusión más detallada en el capítulo 5).

Cuando no se dispone de medidas del tamaño de los conglomerados

Se debe usar un procedimiento ligeramente diferente cuando no se dispone de medidas del tamaño de los conglomerados antes de la selección de la muestra. En este método, todos los conglomerados tendrán la misma probabilidad de selección, o *igualdad de probabilidades*, en lugar de una probabilidad relacionada con su tamaño. Los procedimientos para escoger una muestra de conglomerados con una probabilidad igual se describen en la figura 4-3 y en la 4-4 se ofrece un ejemplo ilustrativo. En ese ejemplo, el objetivo es seleccionar 40 conglomerados de un total de 170.

Figura 4-3: Pasos para seleccionar una muestra aleatoria sistemática de conglomerados con igualdad de probabilidades

- (1) Preparar una lista numerada de sitios o conglomerados, de preferencia, ordenados por región geográfica (por ejemplo, por zonas de una ciudad).
- (2) Calcular el intervalo de muestreo (SI) dividiendo el número total de conglomerados en el dominio (es decir, el grupo destinatario) (M) por el número de conglomerados que se deben seleccionar (a), es decir, $SI = M/a$.
- (3) Seleccionar un número al azar (comienzo al azar o RS) entre 1 y SI. El conglomerado en la lista numerada correspondiente a este número será el de la primera muestra.
- (4) Escoger las unidades subsiguientes agregando el intervalo de muestreo (SI) al número identificado en el paso (3); es decir, $RS + SI$, $RS + SI * 2$, $RS + SI * 3$, etc.
- (5) Seguir este procedimiento hasta acabar con la lista.

Figura 4-4: Ejemplo ilustrativo: selección de una muestra aleatoria sistemática de conglomerados con igualdad de probabilidades

Conglomerado No.	Selección	
001		No. planeado de conglomerados = 40
002	X	Intervalo de muestreo = $170/40 = 4.25$
003		Comienzo al azar entre 1 y $4,25 = 2$
004		Conglomerados seleccionados = 002, 006, 011,
005		
006	X	
007		
008		
009		
010		
011	X	
.		
.		
.		
170 (último)		

Nótese que al seleccionar los conglomerados de muestra, es importante mantener la coma de la fracción decimal en el intervalo de muestreo. Se debe observar la regla siguiente: cuando la fracción decimal del número de selección de la muestra sea menor de 0,5, se escoge el conglomerado con el número menor, y cuando la fracción decimal del número de selección sea de 0,5 o mayor, se escoge el conglomerado con el número mayor. En el ejemplo anterior, el número de selección de la muestra correspondiente al tercer conglomerado de muestra fue de 10,5 y, por consiguiente, se escogió el conglomerado 011 para la muestra.

En este caso se supone también que se debe escoger un número fijo de familias de cada conglomerado de muestra. Sin embargo, en este caso, como la probabilidad de seleccionar un conglomerado no se basó en el número de familias que contiene, el procedimiento lleva a tener elementos de muestra con diferentes probabilidades generales de selección. En otras palabras, la muestra *no tiene un mecanismo de autoponderación*. Esto complicará la situación durante el análisis (véase el capítulo 5).

3.2 Procedimientos de selección de las familias de muestra

En una situación ideal, las familias de muestra deben seleccionarse preparando una lista o un marco de muestreo de todas las familias localizadas dentro de cada conglomerado y escogiendo una muestra de unidades con muestreo aleatorio simple o sistemático. Sin embargo, es probable que la preparación de esas listas de familias sea inaceptablemente costosa y lleve mucho tiempo. Para abreviar, a continuación se describen tres alternativas: el método de *segmentación* y dos variantes del método de *recorrido al azar*.

Método de segmentación

El método de segmentación consiste en dividir los conglomerados de muestra en segmentos más pequeños de tamaño aproximadamente igual, escogiendo al azar un segmento de cada conglomerado y entrevistando a todas las familias en el segmento escogido. El tamaño de cada segmento (es decir, el número de familias en cada uno) debe ser el mismo que el número de familias de muestra que se pretende escoger como objetivo por conglomerado (véase la sección 4 en el capítulo 3). Por ejemplo, si se hubiera determinado que se escogerían 30 conglomerados para una encuesta dada y 50 familias por conglomerado (lo que daría un tamaño de la muestra de $n = 1.500$ familias), el tamaño del segmento escogido como objetivo con el método de segmentación sería de 50 familias.

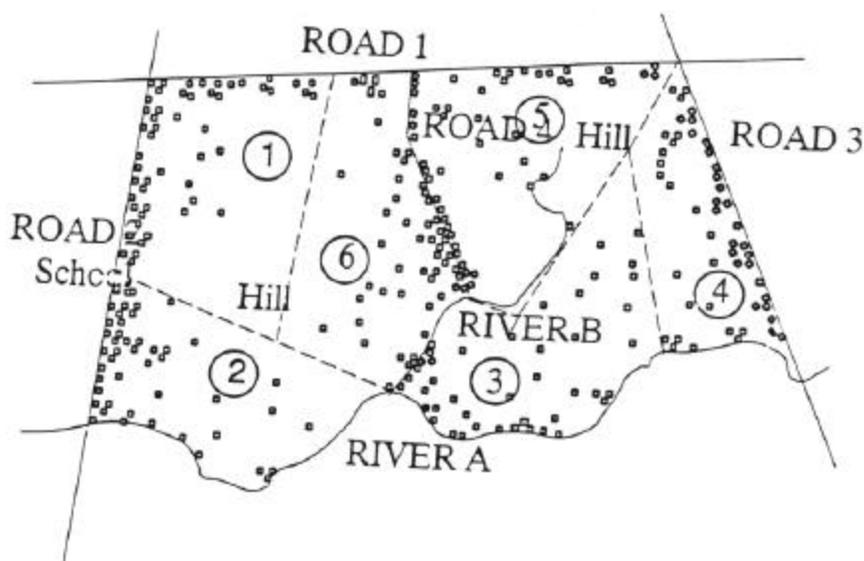
En la figura 4-5 se indican los pasos para usar el método de segmentación y en la figura 4-6 se hace una representación gráfica de un conglomerado hipotético creado con el método de segmentación.

Figura 4-5: Pasos para usar el método de segmentación para escoger las familias de muestra

- (1) Calcular el número de segmentos que se pretende crear. Dividir el número de familias registradas en el último censo por el tamaño del segmento escogido como objetivo. El resultado será el número de segmentos que se piensa crear en el terreno. Por ejemplo, si el último censo indicó que había 250 familias en el conglomerado y el tamaño del segmento escogido como objetivo fue de 40 familias, habría que crear 6 segmentos. (Nótese que al realizar ese cálculo, se deben redondear las fracciones decimales de los segmentos al número entero más próximo.)
- (2) Actualizar el mapa de conglomerados. Empleando un mapa de cada conglomerado, verifique/actualice los límites externos del conglomerado y anote las características internas que puedan ser útiles para dividir el conglomerado en segmentos fácilmente reconocibles.

- (3) Contar el número de familias e indicar su localización en el conglomerado en el mapa. Esa debe ser una operación rápida realizada con el fin de poder dividir el conglomerado en segmentos con un número de familias aproximadamente igual.
- (4) A partir del mapa del conglomerado, dividir el conglomerado en segmentos de igual tamaño. El número de segmentos que se pretende usar es el número determinado en el paso 1 anterior.
- (5) Escoger un segmento al azar.
- (6) Entrevistar a todas las familias localizadas dentro de los límites del segmento escogido al azar.

Figura 4-6: Ejemplo de un conglomerado hipotético dividido en seis segmentos



Road = Camino.

River = Río.

Hill = Colina.

Fuente: Fondo de las Naciones Unidas para la Infancia. 1995. *Monitoring Progress Toward the Goals of the World Summit for Children: A Practical Handbook for Multiple-Indicator Surveys*. Nueva York: UNICEF.

El trabajo sobre el terreno en un conglomerado de muestra determinado en el método de segmentación se considera completo cuando se ha entrevistado a todas las familias del segmento escogido para la encuesta (independientemente del número de sujetos de estudio encontrados en realidad).

Método de recorrido al azar

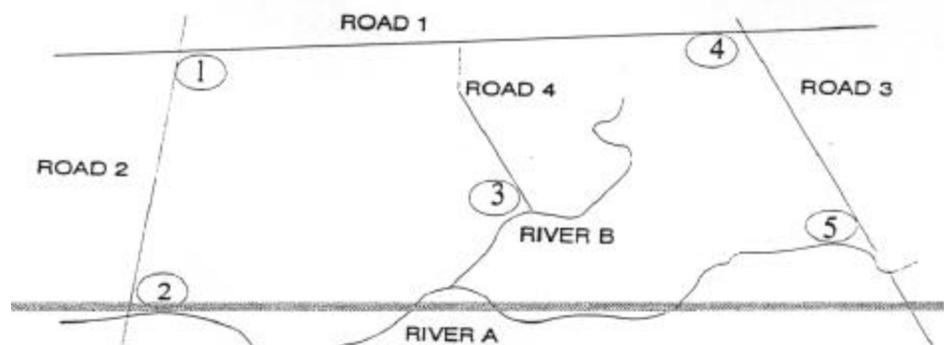
El método de recorrido al azar se usa en encuestas de conglomerados del Programa Ampliado de Inmunización (PAI) y, por lo tanto, es relativamente bien conocido. Entraña (1) la selección aleatoria de un punto de partida y de la dirección en que se hará el viaje dentro de un conglomerado de muestra, (2) la realización de una entrevista a la familia más cercana y (3) la selección continua de la próxima familia más cercana para una entrevista hasta que se haya hecho el número de entrevistas escogido como objetivo.

Este método puede emplearse de una de dos formas y la única diferencia es si hay mapa y, en caso afirmativo, cómo se escoge el punto de partida. A continuación se ofrece un resumen.

Selección del punto de partida en un mapa de los límites del conglomerado

Cuando se dispone de un mapa del conglomerado de muestra, se debe seleccionar un número de posibles puntos de partida en diferentes lugares fácilmente identificables (véase un ejemplo ilustrativo en la figura 4-7) y escoger ahí un punto de partida al azar. La ventaja está en que el personal de supervisión puede escoger el punto de partida antes de que comience el trabajo sobre el terreno, con lo que se reduce cualquier riesgo de sesgo que pudiera surgir cuando se escogen los puntos de partida por conveniencia y no al azar.

Figura 4-7: Mapa de un conglomerado hipotético de muestra con los posibles puntos de partida



Road = Camino.
River = Río.

Fuente: Fondo de las Naciones Unidas para la Infancia. 1995. *Monitoring Progress Toward the Goals of the World Summit for Children: A Practical Handbook for Multiple-Indicator Surveys*. Nueva York: UNICEF.

Selección del punto de partida con el método empleado en el PAI

Para encuestas del PAI, se supone que no hay ningún mapa del conglomerado y que el personal asignado al terreno escogerá un sitio para comenzar. Para ello, seguirá instrucciones que exigen (1) desplazamiento hacia un punto central del conglomerado y selección al azar de la dirección en que se hará el viaje haciendo girar una botella; (2) desplazamiento en línea recta en esa dirección y recuento de todas las familias hasta llegar al borde del conglomerado; y (3) selección aleatoria de un número entre 1 y el número de familias contadas en el punto de partida para la encuesta.

Aunque ninguna de las variantes del método de recorrido al azar exige una medida del tamaño de los conglomerados de la muestra, se debe obtener una estimación del número de familias localizadas en cada conglomerado de muestra para las evaluaciones de los programas del Título II al usar cualquiera de esos dos métodos. Esto permite calcular las probabilidades de muestreo (véase una discusión más detallada en el capítulo 5). El recuento o la estimación del número de familias no tiene que ser excesivamente costoso ni llevar mucho tiempo. En la mayoría de los casos, un informante local con los conocimientos necesarios puede proporcionar una cifra razonable. Si eso no es posible, por lo general, un recorrido rápido por el conglomerado permite proporcionar un recuento de precisión aceptable. Este tipo de procedimiento de *recuento rápido* se usa a menudo en el muestreo por conglomerados.

Si se pretende usar el método de recorrido al azar, habrá que seleccionar una *cuota*, es decir, el trabajador asignado al terreno deberá seguir estableciendo contacto con las familias hasta que se haya localizado un número previamente determinado de sujetos de estudio (por ejemplo, niños menores de cinco años). La cuota debe ser el número de *familias* escogida como objetivo por conglomerado en lugar del número de *sujetos de estudio* necesarios para diferentes indicadores (por ejemplo, niños < 24 meses, niños con un episodio de diarrea en las dos últimas semanas, etc.). Más bien, si el tamaño de la muestra escogida como objetivo por conglomerado es, por ejemplo, de 50 familias, se seguiría el procedimiento de recorrido al azar hasta cuando se haya entrevistado a ese número de familias. Se sigue la trayectoria de las cuotas de familias más que de las cuotas de sujetos de estudio, puesto que sería muy difícil seguir la trayectoria de estos últimos cuando se deben medir varios indicadores. Como se indicó antes, las cuotas no son fijas cuando se usa el método de segmentación; más bien, se considera que el trabajo realizado sobre el terreno en un cierto conglomerado de muestra está completo cuando se ha entrevistado a todas las familias del segmento (independientemente del número de sujetos de estudio encontrado en realidad).

Los tres métodos descritos antes varían mucho en lo que respecta a exposición al riesgo de sesgo. El método de segmentación es el que más se aproxima a una muestra convencional por conglomerados en dos etapas y, por ende, presenta menos posibilidades de crear sesgos. Sin embargo, como es necesario

esbozar un mapa de la comunidad, quizá no sea posible usar este método en todos los medios. El método que ofrece las mejores posibilidades con respecto al sesgo es la variante del método de recorrido al azar en que la familia de la muestra inicial se escoge en un mapa con los límites del conglomerado. El menos preferido es el método de recorrido al azar empleado en el PAI.

3.3 Procedimiento de selección de los sujetos particulares para la encuesta

La recomendación general para la mayoría de las encuestas se centra en la inclusión en la muestra de todos los sujetos de las familias de muestra que reúnan los requisitos establecidos (por ejemplo, niños de ciertos grupos de edad). No obstante, en ciertos tipos de encuesta (por ejemplo, cuando se usan hojas de anotación del consumo de alimentos) y en las comunidades donde las familias ampliadas viven juntas, lo que incluye a todas las mujeres adultas y los niños que forman parte de la muestra, el costo y el gasto de tiempo serían prohibitivos. El procedimiento recomendado en ese caso es la selección al azar de una mujer adulta entre las mujeres de una determinada familia de muestra y la entrevista solamente a esa persona y a su familia e hijos. Esto constituiría un tercer paso en el proceso de selección de la muestra.

Cuando se selecciona uno de varios sujetos posibles dentro de una familia, debe tenerse en cuenta eso en el cómputo de los valores de ponderación de muestreo para poder obtener estimaciones sin sesgo en la encuesta. Esta es una tarea relativamente sencilla: el encuestador solo necesita registrar en el protocolo de la encuesta el número de posibles entrevistados que reúnen los requisitos en cada familia de muestra, de la cual se seleccionó al azar a la persona realmente entrevistada en la encuesta. Eso permite calcular la probabilidad de haber escogido a esa persona. Esa probabilidad se incorpora luego al cálculo de una probabilidad general de muestreo (véase el capítulo 5, figura 5-1 [D]).

3.4 Solución de los problemas operativos de muestreo

Pueden surgir problemas de ejecución aun en las encuestas mejor planeadas. Son típicos de ellos los conglomerados inaccesibles, la falta de respuesta y un número insuficiente de familias en un cierto conglomerado.

Conglomerados inaccesibles

A veces, quizá sea imposible llegar a un conglomerado de muestra, por ejemplo, debido al mal tiempo o a que los caminos están intransitables. Por lo general, lo mejor es reemplazar el conglomerado por otro de características similares escogido al azar. Por ejemplo, si el conglomerado en cuestión está localizado en el extremo norte de la zona del proyecto, debe reemplazarse por otro en la misma zona general, pero que sea *accesible* durante el período del trabajo realizado sobre el terreno con fines de encuesta. Para reducir al mínimo el riesgo de sesgo, los conglomerados de reemplazo deben escogerse

a partir de otros similares; eso no debe crear ningún inconveniente. En la medida de lo posible, el personal de supervisión debe adoptar decisiones sobre los conglomerados de reemplazo.

Falta de respuesta en la encuesta

La falta de respuesta es un problema común en todas las encuestas. Típicamente, se presenta cuando no hay nadie en la casa de las familias de muestra o cuando los sujetos de la encuesta se niegan a dejarse entrevistar. Hay procedimientos establecidos para resolver esos problemas:

No hay nadie en la casa: Cuando no hay respuesta de una familia escogida como objetivo en un conglomerado de muestra, se debe preguntar a los vecinos (1) si la unidad residencial está habitada y, en caso afirmativo, (2) a qué hora suelen estar en casa los residentes. Si la unidad residencial no está ocupada, no se necesita tomar ninguna otra medida. Si lo está, se debe hacer por lo menos otra visita (o mejor, muchas más), de preferencia a la hora del día en que, según declaración del vecino, suelen estar en casa los residentes.

Negación: Cuando los ocupantes de una casa escogida como objetivo se niegan a dejarse entrevistar, tal vez otro miembro o el supervisor del equipo asignado al terreno debe hacer por lo menos otra visita. Sin embargo, se debe dar prioridad a la repetición de las visitas a las personas que no están en casa.

Aunque en el cálculo de los requisitos del tamaño de la muestra es preciso incorporar un cierto grado de falta de respuesta (véase el capítulo 3, sección 3.7), en la medida en que ocurra puede sesgar los resultados de la encuesta. Eso se debe a que a menudo hay diferencias sistemáticas entre las personas que optan por responder y quienes no lo hacen y esas diferencias pueden reflejarse en los indicadores objeto de medición. La mejor forma de abordar ese sesgo potencial de la falta de respuesta es reducir esa falta al mínimo en la medida de lo posible. De conformidad con ello, en los planes de operaciones para el trabajo sobre el terreno se debe dar suficiente tiempo para seguimiento de las personas que no responden. Puesto que en los cálculos del tamaño de la muestra se previó una cierta falta de respuesta, en la mayoría de los casos, el efecto de esta última en el tamaño de la muestra definitivo en la encuesta debe ser tolerable, siempre y cuando no pase mucho del nivel esperado en un gran número de conglomerados.

El número de familias del conglomerado no es suficiente para obtener el tamaño de la muestra escogido como objetivo

Si los cálculos del tamaño de la muestra se han realizado correctamente y se ha tenido en cuenta el tamaño de los conglomerados en los marcos de muestreo disponibles, debe haber suficientes familias para lograr el tamaño de la muestra escogido como objetivo. Además, si se ha incorporado una *reserva* en los cálculos del tamaño de la muestra como se recomendó en el capítulo 3, ya se han compensado los efectos en el tamaño definitivo de la muestra de encuesta. En caso de que se presente esa situación, se debe desaconsejar a los equipos asignados al terreno que escojan otras familias de los

conglomerados vecinos. Más bien, deben concentrar sus esfuerzos en reducir al mínimo el número de familias que no responden y luego pasar al siguiente conglomerado que se les haya asignado.

3.5 Procedimiento de selección de las muestras en las zonas de comparación

Por lo general, se espera que los grupos de comparación estén formados por poblaciones de uno o más distritos, municipios u otras unidades administrativas de los alrededores con características similares a las del programa objeto de evaluación. El proceso de selección consta normalmente de dos etapas. La primera tiene que ver con la identificación de grupos que se ciñen a los criterios de similitud. La selección puede hacerse *deliberadamente* (es decir, las características del grupo podrían definirse con anterioridad y la selección podría hacerse de acuerdo con los criterios acordados), a menos que varias zonas tengan un perfil similar al de la zona del programa, caso en el cual se podría escoger una al azar. (Véase una orientación más detallada en la Guía de seguimiento y evaluación.) Una vez que se haya definido el universo de encuesta de la zona de comparación, solo resta seleccionar una muestra de conglomerados y familias representativos de dicha zona. Los procedimientos de muestreo son idénticos a los ya descritos con respecto a las encuestas de la población en general.

En figura 4-8 se ofrece un ejemplo ilustrativo de las decisiones referentes al muestreo en un grupo de comparación.

Figura 4-8: Ejemplo ilustrativo del diseño y la selección de muestras en una zona de comparación

Supongamos que se estuviera realizando un programa centrado en mejorar la producción y elaboración agrícolas en dos distritos de un país. Se ha empleado un método de evaluación con una prueba *a priori* y otra *a posteriori*, con el diseño del grupo de comparación para fines de la evaluación del programa. Para la zona de comparación, todos los distritos adyacentes a los dos distritos del proyecto se consideran como posibles candidatos. Dos de ellos se eliminan porque tienen características socioeconómicas diferentes de las de los distritos del proyecto y otro porque es el objetivo de un programa comparable financiado por otro donante. Del resto de los distritos adyacentes, se escogen los dos más similares a los del proyecto.

En cada distrito donde haya grupos de comparación, se escoge una muestra de 30 conglomerados con un procedimiento de selección aleatoria sistemática con PPS y 40 familias de cada conglomerado de muestra empleando el método de segmentación en dos rondas de la encuesta : una encuesta básica y otra de seguimiento.

4. Procedimientos de selección de las muestras para programas con cobertura limitada de la población

4.1 El problema del muestreo

Algunos programas del Título II tienen un grado bastante limitado de cobertura de la población en general. Por ejemplo, la población posiblemente beneficiaria de un programa de crédito para madres solteras con hijos menores de cinco años podría ascender a menos de 5% de las familias en la zona geográfica escogida para un programa. En esos casos, no es apropiado tratar de medir el efecto del programa en la población en general; sería casi imposible detectar cambios en los indicadores en ese nivel, aun cuando hayan ocurrido cambios considerables para los beneficiarios del proyecto. Como resultado, las actividades de evaluación del programa se destinan a menudo a evaluar los cambios en los indicadores solamente entre las familias beneficiarias. A continuación se presentan otras tres estrategias de muestreo para esas situaciones: restricción del universo de la encuesta; detección sistemática; y uso de los marcos de las listas.

4.2 Otras estrategias de muestreo

Restricción del universo de la encuesta

Hay un método apropiado cuando los beneficiarios del proyecto están muy concentrados en ciertas partes de la zona del proyecto (por ejemplo, en municipios o poblados específicos). En esos casos, el universo de la encuesta se puede limitar a los puntos de impacto del programa y el marco de muestreo, a ese subconjunto de conglomerados. Con esa restricción, las estrategias de muestreo descritas antes podrían emplearse sin otras modificaciones. Este método no se aplicaría en la mayoría de los casos, ya que, por lo general, los beneficiarios del programa están dispersos en toda la zona geográfica escogida como objetivo por un proyecto determinado.

Detección sistemática

Cuando los beneficiarios se encuentran dispersos en una región geográfica por toda la zona del proyecto, un método de uso común consiste en introducir un *procedimiento de detección sistemática*. Al comienzo de la operación de la encuesta realizada sobre el terreno, se pregunta a las familias que pueden formar parte de la muestra si se han beneficiado del proyecto. (Se podría agregar esa clase de preguntas al comienzo del cuestionario.) Los trabajadores asignados al terreno realizan la entrevista de la encuesta solamente cuando se trata de familias beneficiarias del proyecto, pasando por alto a quienes no lo sean. De lo contrario, se puede acopiar una pequeña cantidad de información de las familias no beneficiarias para poder hacer comparaciones entre las beneficiarias y otras. Este método serviría para asegurarse de que el gasto de recursos se realice con acopio de información sobre los beneficiarios del proyecto.

Esta estrategia exigiría que se agregaran familias al tamaño de la muestra escogido como objetivo. Por ejemplo, si se creyera que solo alrededor de 20% de las familias en una zona geográfica escogida como objetivo serían beneficiarias del proyecto, únicamente una de cinco familias reuniría los requisitos para llenar los cuestionarios. En ese caso, habría que quintuplicar el número de familias de la muestra.

Para ilustrar eso, supongamos que se determinara que los requisitos del tamaño de la muestra para una encuesta dada fueran de $n = 1.500$ familias y que el plan de muestreo fuera tomar 30 conglomerados de 50 familias cada uno. Si se creyera que solamente 20% de las familias de la zona objetivo del proyecto fueran beneficiarias de éste, habría que ponerse en contacto con un total de $n = 7.500$ ($7.500 = 1.500 / 0,20$) familias. Si se mantuviera la meta de 30 conglomerados, eso exigiría ponerse en contacto con $n = 250$ familias por conglomerado. En muchos casos, eso constituiría todo el conglomerado, caso en el cual la estrategia de muestreo sería entrevistar a todas las familias en los conglomerados de muestra. Sin embargo, en muchos medios, los conglomerados en marcos de muestreo disponibles (por ejemplo, zonas de enumeración del censo u otras unidades administrativas) son menores de 250 familias, lo que exige selección de un mayor número de conglomerados. Por ejemplo, si el tamaño medio de conglomerados en un lugar dado fuera de 100 familias, sería necesario escoger una muestra de 75 conglomerados y comunicarse con todas las familias localizadas dentro de esos conglomerados ($7.500/100 = 75$ conglomerados).

Uso de marcos de listas

Donde se dispone de listas de beneficiarios del programa, éstas ofrecen una manera mucho más eficiente de escoger una muestra de familias beneficiarias que la detección sistemática. En este caso se pueden emplear dos métodos. Uno consiste en escoger una muestra de familias beneficiarias con un muestreo aleatorio simple o sistemático. Sin embargo, esto podría dar origen a una operación logísticamente difícil sobre el terreno, puesto que las familias de la muestra podrían estar muy dispersas en un gran número de pueblos. Por otra parte, se podría agrupar a los beneficiarios de la lista por región geográfica y realizar el muestreo de los conglomerados resultantes de familias beneficiarias. El marco resultante de muestreo por conglomerados se consideraría exactamente lo mismo que un marco de muestreo por conglomerados más convencionales (es decir, zonas de enumeración del censo). Entonces, podrían emplearse los procedimientos de muestreo para encuestas de la población en general descritos antes.

5. Cuestiones de diseño para las encuestas de seguimiento

5.1 *Retención o reemplazo de conglomerados de muestra*

Una de las cuestiones clave en materia de diseño en cualquier encuesta de varias rondas en que se usa el muestreo por conglomerados es si se mantienen los mismos conglomerados en cada encuesta o se escoge una nueva muestra. Desde el punto de vista estadístico, la estrategia preferida es mantener los

mismos conglomerados. Eso se debe a que las características generales y los patrones de comportamiento de las personas de zonas geográficas pequeñas tienden a estar correlacionados con el tiempo. El efecto de esta correlación es reducir hasta cierto punto la variabilidad de las estimaciones de cambio hechas en la encuesta, lo que facilita un poco la tarea de detectar el cambio *real*. Aunque la magnitud del aumento esperado en la eficiencia del muestreo varía entre medios e indicadores, casi siempre habrá algún aumento de la precisión de las mediciones. Por ende, la recomendación general es mantener la misma muestra de conglomerados y, al mismo tiempo, escoger una nueva muestra de familias en cada conglomerado siguiendo uno de los procedimientos descritos antes.

Esta estrategia acarrea un grave riesgo: la tentación de asignar recursos del programa desproporcionadamente a favor de las zonas que, según se sabe, son el punto de enfoque de la evaluación del programa. Si el programa concentra los recursos de esa forma, esas zonas pueden mostrar resultados más positivos de los que se habrían obtenido en toda la zona del programa en general. Los administradores del programa deben protegerse de desviar recursos hacia las zonas escogidas como objetivo para evaluación. Si los evaluadores temen que ha ocurrido eso, lo más seguro sería escoger una nueva muestra de conglomerados para la ronda de la encuesta de seguimiento. Eso podría reducir algo la eficiencia del muestreo para la evaluación del programa, pero también disminuiría el riesgo de sesgo, que es una consideración más importante dentro del marco ampliado de las cosas.

5.2 Efectos de los cambios en los métodos de encuesta

Otro problema está en que los efectos *reales* del programa se pueden distorsionar o *confundir* por causa de cambios en otros factores entre una ronda y la siguiente. Los *factores de confusión* pueden dividirse aproximadamente en dos categorías amplias: externos e internos.

Los *factores de confusión externos* son los cambios en los factores que obran en la población objeto de estudio sobre los cuales el control ejercido por los programas y evaluadores es poco o nulo. Entre los ejemplos cabe citar cambios en la composición de la población por causa de emigración, cambios del medio económico, desastres naturales, etc. En las evaluaciones de programas se suele lidiar con esos factores mediante el uso de zonas y grupos testigos, que ayudan a separar los efectos *reales* del programa de los causados por otros factores no relacionados con el mismo, y métodos estadísticos multifactoriales durante el análisis de datos.

Los *factores de confusión internos* son los cambios en los factores internos de una actividad de evaluación de programas, controlados por el personal asignado al terreno encargado de realizar la evaluación. Son ejemplos de ellos los cambios en la definición del universo de la encuesta, la redacción del cuestionario, el método de muestreo y la calidad o el momento de realización (con respecto a la periodicidad) del trabajo de la encuesta realizado sobre el terreno. En particular cuando se usa una muestra de un tamaño relativamente pequeño, los resultados de la evaluación pueden ser bastante sensibles a cambios aun modestos en los métodos o en la ejecución de la encuesta. En este caso la lección está en que los métodos y los resultados de la ejecución de la encuesta deben ser tan constantes como sea posible.

5.3 Destrucción de las muestras

A veces, los planes de usar la misma muestra de conglomerados en las encuestas básica y de seguimiento deben abandonarse por causa de desastres naturales, problemas de seguridad, etc. En esos casos, la mejor línea de acción es seguir el procedimiento recomendado para conglomerados inaccesibles y reemplazar el conglomerado sobre el que no es posible obtener datos con otro de características similares escogido al azar (véase la sección 3.4 precedente). El uso de la selección al azar reducirá al mínimo el riesgo de sesgo. La preferencia por características similares tiene por fin tratar de limitar la variabilidad del muestreo entre las distintas rondas de la encuesta. El personal de supervisión debe tomar las decisiones sobre la selección de conglomerados de reemplazo.

Sin embargo, si la falta de acceso al conglomerado es solamente temporal al programar la encuesta de seguimiento, las personas encargadas de realizarla deben esperar hasta que el conglomerado esté accesible de nuevo en lugar de seleccionar un reemplazo, con la condición de que eso no cause una demora inaceptable del trabajo de encuesta realizado sobre el terreno

5

Análisis de los datos

1. Visión panorámica

Una vez que se haya acopiado la información de la encuesta para el seguimiento y la evaluación de un programa e ingresado a la base de datos del computador, solo resta analizarla. Esto entrañará el cálculo de los diversos indicadores medidos en las encuestas y la evaluación de la magnitud y significación estadística de los cambios con el tiempo y las diferencias entre los grupos de comparación. En la guía de seguimiento y evaluación de esta misma serie se ofrecen directrices para analizar los datos para la evaluación de programas del Título II. En esta sección se presta atención únicamente a dos cuestiones relacionadas con el análisis, en las que influye mucho la forma de realizar el muestreo: *los valores de ponderación del muestreo y el cálculo de los errores estándar de las estimaciones de la encuesta.*

2. Ponderación de los datos

Solamente uno de los planes de muestreo por conglomerados descrito en el capítulo anterior producirá muestras con un mecanismo de autoponderación, es decir, en que los conglomerados de muestra se escogen con probabilidad proporcional al tamaño (PPS) y el método de segmentación se usa para seleccionar a las familias de la muestra. Los demás planes producirán muestras sin un mecanismo de autoponderación o muestras en que las familias tienen distintas probabilidades de selección. Esto debe compensarse en la etapa de análisis de datos. Si no se hace eso, habrá un sesgo de estimación.

Los valores de ponderación compensan la desigualdad de probabilidades de selección. El método estándar para corregir esa desigualdad consiste en aplicar los valores de ponderación del muestreo a los datos de la encuesta durante el análisis, multiplicando el valor del indicador por el valor de ponderación. La ponderación apropiada de muestreo de cada sujeto de la muestra es sencillamente la cantidad recíproca de la probabilidad de selección de ese sujeto o el inverso de la probabilidad.

Ecuación básica 3: Proporción

$$W_i = 1/P_i$$

CLAVE:

W_i = ponderación del muestreo de los elementos en el i° conglomerado y

P_i = probabilidad de selección de los elementos en el i° conglomerado.

Para calcular *los valores de ponderación* del muestreo, hay que calcular la probabilidad de selección (P_i). La fórmula dependerá de la variante usada del diseño de la muestra por conglomerados de dos (o tres) etapas. La figura 4-1 muestra los cálculos de la probabilidad de muestreo en cuatro de las posibles combinaciones: (A) PPS (primera etapa)/segmentación (segunda etapa); (B) PPS/recorrido al azar; (C) igualdad de probabilidades/recorrido al azar; y (D) PPS/segmentación, pero con un entrevistado. (Las opciones no presentadas incluyen igualdad de probabilidades/segmentación; igualdad de probabilidades/recorrido al azar, pero con un entrevistado; e igualdad de probabilidades/segmentación, pero con un entrevistado. Los analistas pueden adaptar las opciones para ajustarse a estos otros casos.) Nunca se necesitará el cálculo del ejemplo A (probabilidad proporcional al tamaño/segmentación) puesto que, según se ha demostrado, este diseño produce una muestra con un mecanismo de autoponderación.

Figura 5-1: Procedimientos para calcular las probabilidades de muestreo de los elementos de la muestra (P_i) con diseños selectos de muestreo por conglomerados en dos etapas

A. PPS en la primera etapa, método de segmentación en la segunda:

$$P_i = (m * M_i/M) * 1/S_i = m * C/M$$

CLAVE

- m = número de conglomerados de muestra escogidos
- M_i = medida del tamaño para el i° conglomerado
- M = medida total del tamaño para el universo de la encuesta
($M = \sum M_i$)
- S_i = número de segmentos creados en el i° conglomerado
- C = tamaño estándar (es decir, constante) del segmento

Nótese que como este diseño produce una muestra con un mecanismo de autoponderación, no se exige la aplicación de valores de ponderación del muestreo durante el análisis.

Figura 5-1: Procedimientos para calcular las probabilidades de muestreo de los elementos de la muestra (P_i) con diseños selectos de muestreo por conglomerados en dos etapas (continuación).

B. PPS en la primera etapa, número constante de elementos escogido en la segunda etapa con el método de recorrido al azar:

$$P_i = (m * M_i/M) * k/N_i$$

CLAVE:

m	=	número de conglomerados de muestra escogidos
M_i	=	medida del tamaño para el i° conglomerado tomada del marco de muestreo
M	=	medida total del tamaño para el universo de la encuesta ($M = \sum M_i$)
k	=	número constante de familias escogidas por conglomerado
N_i	=	número total de familias en el i° conglomerado, es decir, M_i actualizado sobre el terreno con el recuento estimado o real de familias

C. Igualdad de probabilidades en la primera etapa, número constante de elementos escogidos en la segunda etapa con el método de recorrido al azar:

$$P_i = (1/m) * (k/N_i)$$

CLAVE:

m	=	número de conglomerados de muestra escogidos
k	=	número constante de elementos de muestra escogidos por conglomerado
N_i	=	número total de familias en el i° conglomerado

D. Diseño (B) anterior, pero con un entrevistado escogido por familia de muestra (por ejemplo, una mujer adulta)

$$P_i = (m * M_i/M) * (k/N_i) * 1/R_{ij}$$

CLAVE:

m	=	número de conglomerados de muestra escogidos
M_i	=	medida del tamaño para el i° conglomerado
M	=	medida total del tamaño para el universo de la encuesta ($M = \sum M_i$)
k	=	número constante de familias escogidas por conglomerado
N_i	=	número total de familias en el i° conglomerado
R_{ij}	=	número total de entrevistados que reúnen los requisitos establecidos en la j^{a} familia

Surgirá un problema cuando se apliquen los valores de ponderación del muestreo al análisis de datos realizado con programas estándar de informática (por ejemplo, Epi-Info, SPSS). La aplicación de los valores de ponderación a esos programas inflará el número de casos de muestra y, por consiguiente, implicará un tamaño de la muestra mayor del real.³ Como resultado, las pruebas estadísticas de las diferencias y los cambios con el tiempo se basarán en un tamaño incorrecto de la muestra, lo que podría llevar a conclusiones erróneas sobre los efectos de los programas. Por ejemplo, los cambios o las diferencias entre los grupos de comparación sin significación estadística a partir del tamaño real de la muestra parecerán ser significativos sobre la base del número ponderado de casos.

En compensación, se emplean a menudo los *valores de ponderación normalizados*. Éstos asignan un valor de ponderación a cada observación de la muestra que refleja su probabilidad relativa de selección en comparación con otras observaciones, pero no cambian el tamaño de la muestra en la encuesta general. Los valores de ponderación normalizados (w_i') para elementos de la muestra en el i^o conglomerado se calculan de la manera siguiente:

$$w_i' = w_i n_i / \sum w_i n_i$$

Puesto que cada elemento de un cierto conglomerado tiene la misma probabilidad de selección, cada uno también recibirá la misma ponderación normalizada. La figura 5-2 ilustra el cómputo de los valores de ponderación normalizados con datos hipotéticos de la encuesta.

Con el fin de hacer uso de los valores normalizados de ponderación del muestreo durante el análisis de datos, habrá que incluir una variable apropiada de ponderación en el archivo de datos de la encuesta que se pretende analizar. Los valores de ponderación normalizados podrían calcularse a mano o con una hoja de cálculo estadístico y anotarse como variable durante el ingreso de datos. De lo contrario se podrían ingresar las probabilidades de selección para las etapas primera y segunda y calcular los valores de ponderación con instrucciones apropiadas de los programas SPSS e Epi-Info.

3. Cuando se multiplica el valor de una variable o de un indicador por la ponderación del muestreo, el resultado es la *observación aparente* de que el tamaño de la muestra es mayor del real.

Figura 5-2: Ejemplo ilustrativo del cómputo de las probabilidades de selección, los valores de ponderación y los valores de ponderación normalizados del muestreo: datos hipotéticos

En este ejemplo se presenta el cálculo de los valores de ponderación normalizados para los cinco primeros conglomerados de una muestra escogidos en una encuesta hipotética.

CLAVE:

n_i = el número de elementos de muestra escogidos en un conglomerado i (que fue siempre de 50, puesto que se escogió un número constante de elementos por conglomerado)

P_i = probabilidad general de selección de los elementos de muestra en el conglomerado i

w_i = ponderación del muestreo de los elementos de muestra en el conglomerado i

w_i' = ponderación normalizada del muestreo de los elementos de muestra en el conglomerado i

No. del conglomerado

	n_i	P_i	w_i	$w_i n_i$	w_i'
1	50	0,033	30,30	1515,00	0,0167
2	50	0,022	45,45	2272,50	0,0251
3	50	0,0030	33,33	1666,50	0,0184
4	50	0,043	23,26	1163,00	0,0125
5	50	0,023	43,48	2174,00	0,0240
.					
.					
.					
Total	2.500			90.526,28	

3. Estimación de los errores estándar

La prueba de significación estadística de los cambios o tendencias observados exige estimaciones de la magnitud del error de muestreo relacionado con las estimaciones de la encuesta, que se llaman comúnmente *errores estándar*. La estimación del error de muestreo puede ser una tarea bastante compleja, según el diseño de la muestra empleado para recopilar los datos correspondientes. A medida que aumenta la complejidad del diseño de una muestra (por ejemplo, cuando se usa estratificación, muestreo por conglomerados y varias etapas de selección de muestras), se complican bastante los procedimientos para estimar los errores estándar. La estimación de los errores estándar de esos diseños traspasa los límites de los problemas que se pueden resolver razonablemente en muchas

evaluaciones de programas del Título II sin asistencia de un estadígrafo..

Por desgracia, los programas estadísticos de informática comunes, como SPSS y Epi-Info, no ofrecen una solución adecuada a este problema. Aunque ambos permiten estimar los errores estándar de los cambios observados y de las diferencias entre los grupos de comparación en cuanto a indicadores y realizar pruebas estadísticas apropiadas, en los errores estándar producidos por esos programas de informática se supone el empleo del muestreo aleatorio simple para acopiar datos de la encuesta. Como es muy probable que se emplee el muestreo por conglomerados en las encuestas de evaluación de programas del Título II, por lo general, se subestimarán los errores estándar estimados producidos por esos programas de informática. Como resultado, algunos cambios o tendencias se juzgan como cambios *reales* cuando, de hecho, son demasiado pequeños para ser estadísticamente significativos..

A falta de programas de informática de mayor alcance para resolver el problema, una opción sería compensar la subestimación prevista de los errores estándar estableciendo criterios más estrictos para juzgar la significación estadística. Por ejemplo, en lugar de usar $p < 0,05$ como el límite para juzgar como significativo un cambio o una diferencia observados en un indicador, se podría usar $p < 0,04$ o aun $< 0,03$. De esa forma, se reduce el peligro de juzgar incorrectamente un cambio observado como significativo, sin tener que aumentar la complejidad de los análisis estadísticos de los datos de encuesta ni usar programas de informática en los cuales no ha sido adiestrado ni tiene experiencia previa el personal de la localidad. En definitiva, cuando haya que hacer análisis cuidadosos de los datos de evaluación de los programas, se necesitarán programas de informática (como STATA o SUDAAN) que permitan trabajar con diseños de muestreo complejos.

Apéndice 1

Lista de indicadores genéricos de programas del Título II

Categoría	Nivel	Indicador
Salud, nutrición y atención materno-infantil	Impacto	% de niños de 24 a 59 meses con retraso del crecimiento (puntuación Z de la talla para la edad)
		% de niños con insuficiencia de peso por grupo de edad (puntuación Z del peso para la edad)
		% de lactantes amamantados en las 8 horas siguientes al nacimiento
		% de lactantes menores de 6 meses alimentados exclusivamente con leche materna
		% de lactantes de 6 a 10 meses que recibieron alimentos complementarios
		% de lactantes que recibieron alimentación continua durante un episodio de diarrea
		% de lactantes que recibieron alimentos extra por 2 semanas después de un episodio de diarrea
	Seguimiento anual	% de niños que reúnen los requisitos para el programa de observación y promoción del crecimiento
		% de niños inmunizados contra el sarampión a los 12 meses
		% de comunidades con organización de salud en su medio
% de niños en el programa de promoción del crecimiento c/ aumento de peso en los 3 últimos meses		
Agua y Saneamiento	Impacto	% de lactantes con un episodio de diarrea en las dos últimas semanas
		Litros de uso de agua en la familia por persona
		% de la población con buenos hábitos de lavado de las manos
		% de familias con acceso a servicios adecuados de saneamiento (también seguimiento anual)
	Seguimiento anual	% de familias con acceso a agua apta para consumo todo el año
		% de instalaciones de abastecimiento de agua y saneamiento por comunidad
Consumo de alimentos de la familia	Impacto	% de familias que consumen la cantidad mínima diaria de los alimentos necesarios
		Número de comidas/refrigerios diarios
		Número de porciones de diferentes clases y grupos de alimentos consumidos
Productividad agrícola	Impacto	Rendimiento anual de los cultivos escogidos como objetivo
		Deficiencias de rendimiento (reales frente a potenciales)
		Variabilidad del rendimiento en diversas condiciones
		Valor de la producción agrícola por familia vulnerable
		Meses de provisiones de cereales de la familia
		% de cultivos perdidos por causa de plagas o del medio ambiente
	Seguimiento anual	Rendimiento actual de los cultivos escogidos como objetivo
		Número de hectáreas en que se adoptaron prácticas mejoradas
		Número de instalaciones de almacenamiento construidas y en uso
Ordenación de recursos naturales	Impacto	Erosión imputada del suelo
		Fertilidad imputada del suelo
		Rendimientos o variabilidad de los rendimientos (también seguimiento anual)
	Seguimiento anual	Número de hectáreas en que se siguieron prácticas de ordenación de recursos naturales
Tasa de supervivencia de plántulas y árboles jóvenes (brinzales)		
Caminos*	Impacto	Margen de precios de los insumos agrícolas de una zona a otra
		Disponibilidad de insumos agrícolas de importancia
		Costo del transporte de los productos alimentarios básicos por estaciones
		Volumen de productos agrícolas transportados por las familias al mercado
		Volumen de tráfico por tipo de vehículo
	Seguimiento anual	Kilómetros de caminos rehabilitados de la finca al mercado
		Medidas anuales selectas de los indicadores de impacto

* Construidos como parte de programas de suministro de alimentos o de otros productos por trabajo.

Apéndice 2

Tamaño de la muestra necesario para determinadas combinaciones de P_1 y cambios o diferencias por detectarse entre los grupos de comparación
($\alpha = 0,95$ y $\beta = 0,90$)

Cambio/diferencia por detectarse (P_1-P_2)												
	Aumento						Reducción					
P_1	0,05	0,10	0,15	0,20	0,25	0,30	0,05	0,10	0,15	0,20	0,25	0,30
0,10	1491	428	211	129	87	63	942	154	29	•	•	•
0,15	1970	540	257	152	101	71	1491	300	97	32	5	•
0,20	2382	634	295	171	112	78	1970	428	158	69	29	10
0,25	2724	711	326	186	120	83	2382	540	211	101	51	26
0,30	2999	771	348	197	125	86	2724	634	257	129	71	40
0,35	3204	814	364	203	128	87	2999	711	295	152	87	52
0,40	3341	840	371	206	128	86	3204	771	326	171	101	63
0,45	3410	848	371	203	125	83	3341	814	348	186	112	71
0,50	3410	840	364	197	120	78	3410	840	364	197	120	78
0,55	3341	814	348	186	112	71	3410	848	371	203	125	83
0,60	3204	771	326	171	101	63	3341	840	371	206	128	86
0,65	2999	711	295	152	87	52	3204	814	364	203	128	87
0,70	2724	634	257	129	71	40	2999	771	348	197	125	86
0,75	2382	540	211	101	51	26	2724	711	326	186	120	83
0,80	1970	428	158	69	29	10	2382	634	295	171	112	78
0,85	1491	300	97	32	5	•	1970	540	257	152	101	71
0,90	942	154	29	•	•	•	1491	428	211	129	87	63

Apéndice 3

¿Cuáles son la confianza y el poder estadísticos de los subgrupos?

Problema: El tamaño de la muestra es n computado sobre la base de un grupo destinatario (por ejemplo, niños menores de 23 meses), pero ¿cuáles son la confianza (α) y el poder (β) estadísticos de otros subgrupos (por ejemplo, niños de 12 a 23 meses)?

Suposiciones: $\alpha = 0,95$; $\beta = 0,80$;

La magnitud de la diferencia, $p_2 - p_1$, es igual en ambos subgrupos

- El poder y la significación estadísticos de cualquier subgrupo mayor (n es mayor) que el grupo original serán mayores y aceptables por definición.
- Si un subgrupo tiene un valor pequeño de n , los valores aproximados son:

Índice del tamaño de la muestra	α	β
0,9 (10% menor)	0,937	0,776
0,8	0,920	0,748
0,75	0,911	0,732
0,6	0,871	0,678
0,5 (la mitad)	0,834	0,634
0,33 (un tercio)	0,740	0,540
0,25 (un cuarto)	0,673	0,478
0,20 (un quinto)	0,619	0,433
0,17 (un sexto)	0,576	0,399
0,14 (un séptimo)	0,541	0,372
0,13 (un octavo)	0,513	0,349
0,10 (un décimo)	0,465	0,314

El poder y la significación estadísticos son demasiado imprecisos para sacar conclusiones lógicas con respecto a los subgrupos si el índice del tamaño de la muestra es de la mitad o menos.

Preparado por Bob Magnani y Tony Turner para el Taller de Muestreo para la Gestión de la Ayuda Alimentaria, 1999. (www.foodaid.org)

Food and Nutrition Technical Assistance Project
Academy for Educational Development
1825 Connecticut Avenue, NW
Washington, DC, 20009-5721.

Tel.: 202-884-8000 Fax: 202-884-8432

Correo electrónico: fanta@aed.org

Internet: <http://www.fantaproject.org>